Characterising functions of long non-coding RNAs in Drosophila embryogenesis

Jun Wen Eugene Seah

Girton College

Cancer Research UK Cambridge Institute

University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

May 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. The length of this dissertation does not exceed the word limit of 60,000 words as specified by the University of Cambridge Degree Committee.

Jun Wen Eugene Seah

May 2019

Abstract

An appreciation of the complexity of the mammalian transcriptome has expanded our understanding of eukaryotic genome regulation with the discoveries of functional non-coding ribonucleic acids (RNAs). In recent years, the number of studies in the field of long non-coding RNA (lncRNA) biology has increased dramatically. Transcriptomic analyses of the eukaryotic genome revealed that the genome is pervasively transcribed and contains a vast number of lncRNA transcripts, most with unknown functions. Although relatively little is known about lncRNAs in general, a few have been shown to function in the regulation of gene expression during development and have been associated with a number of diseases. The aim of my dissertation is to investigate the impact of lncRNAs loss of function during embryogenesis of *Drosophila* melanogaster. I chose this model organism for its well documented developmental stages and the plethora of established tools available to facilitate genetic studies. Twenty-two lncRNA candidates were chosen based on their conservation at the sequence level and similar expression profiles across 5 *Drosophila* species, suggesting their potential for biological importance. The CRISPR/Cas9 system was used to generate lncRNA mutants and their requirement for development and the phenotypic consequences of losing each lncRNA was investigated. 13 lncRNA mutants were generated and two of them, lncRNA-9 and lncRNA-3 respectively, were required for viability, with homozygous mutants showing full lethality. Majority of the lncRNA-9 *null* mutant embryos were found to be unable to complete embryonic development and whereas lncRNA-3 null mutants had a pupal lethal phenotype. None of the lncRNA mutants were found to be required for fertility. I characterised the sub-cellular localization of lncRNA-9 during embryogenesis using a combination of RNA Fluorescence In Situ Hybridization (RNA-FISH) and Immunofluorescence (IF) approaches. An analysis of the transcriptome of lncRNA-9 mutants, in comparison to controls, was carried out to discover the genes that were mis-regulated and responsible for the observed lethality. Our investigation of lncRNA-9 revealed a nuclear lncRNA that is expressed in neurons and preliminary results from GO analysis revealed a loss of lncRNA-9 resulted in a reduction in the expression of neuronal genes involved in chemical synaptic transmission activities. Further characterization of lncRNA-9 will allow us to understand how lncRNAs contributes to various neural circuits and the overall signalling in the *Drosophila* connectome.

Acknowledgements

This PhD journey has been a long one spanning 4 years and 7 months and it is one that has been filled with lots of ups and downs. Starting at Cold Spring Harbour Laboratory, New York in 2013 as a visiting research student after graduating from Nanyang Technological University, I would not have foreseen myself being part of this exciting adventure and moving to the University of Cambridge in 2014 to start my PhD studies. Throughout this journey, I have interacted with many people from all walks that have made significant impact in my life.

The first person that I would like to acknowledge is Greg. Without him, I would not have the opportunity to set foot for the first time at two of the greatest research institutions across two continents. He has given me the chance to work on a really challenging topic which has shaped the way I think about biology. Through the various conversations with him at work, he has helped me shaped my project and also given me the confidence to speak up during my presentations. Outside of work, Greg has an amazing “fatherly” character and he will always take care of us at social events and retreats. I am truly grateful for the mentorship and dedication that he has provided and the sense of belonging in this family.

Many lab members have key roles throughout my project and I will like to express my gratitude to you. Ben, who got me into working on this topic and taught me how to work with *Drosophila*. Nic and Joaquina, from whom I have learnt a great deal about data analysis and always keeping me motivated and excited about life and lncRNAs. Ashley and Ilaria, they were always there for me when I need them and ensuring that I am well-taken care of in the lab. A big thank you to everyone in the Hannon lab for the support and wonderful time that we have shared.

To my family back in Singapore, thank you for being patient and supportive of my studies. To my friends from Girton college, you have taught me a lot about your area of specialties and also shared a slice of your life with me. Thank you for all of these wonderful memories.

Table of Contents

List of Figures

Supplementary Figures:

List of Tables

Supplementary Tables:

List of abbreviations

° C ..... degree Celsius
bp..... base pair(s)
cm..... centimetre
dNTP ..... deoxynucleotide
DMSO ..... dimethyl sulfoxide
dsRNA..... double stranded RNA
EDTA ..... ethylenediaminetetraacetic acid
E. coli ..... Escherichia coli
fc ..... fold change
hr ..... hour(s)
IRES ..... internal ribosomal entry site
Kb..... kilo base(s)
MgCl2 ..... magnesium chloride
M ..... molar
µg ..... microgram
µl ..... microliter
µM ..... micromolar
ml ..... milliliter
mM ..... millimolar
mins ..... minute(s)
ng..... nanogram
nt ..... nucleotide(s)
PCR ..... polymerase chain reaction
RNP ..... ribonucleoprotein
rpm ..... revolutions per minute
NaCl ..... sodium chloride
sec ..... second(s)
Tris ..... tris(hydroxymethyl)aminomethane
UTR..... untranslated region
V..... voltage

Chapter 1 |

Introduction

1.1 Regulatory Ribonucleic acid: a brief history of the RNA world

Understanding the regulatory programs that are involved in coordinating various cellular activities has been a central theme in molecular biology since the elucidation of the double helical structure of deoxyribonucleic acid (DNA) in the 1950s (Watson and Crick, 1953; Franklin and Gosling, 1953; Langridge *et al.,* 1960). Since then, there have been many advances towards understanding the processes that translate genotype into phenotype. In 1957, the lecture on protein synthesis and functions of the gene by Nobel laureate Francis Crick had grounded our understanding of how life begins (Crick, 1958). The genetic information of every organism was considered to be defined by the genetic code and determine the messenger RNA (mRNA) transcript that is coded and the amino acid sequence of a protein. The central dogma of gene expression that Crick had proposed made such a lasting impact on our understanding of the connection between nucleic acid and proteins that it is considered to be the "rule" even today (Figure 1). Together with the concept of one gene-one protein articulated by Beadle and Tatum, this framework became the cornerstone in every aspect of molecular biology research and teaching (Beadle and Tatum, 1941).

For many years, proteins were recognised to be the final effector of many regulatory functions in the cell, and the production of dysfunctional proteins contributed to developmental abnormalities and diseases (Scheper *et al.,* 2007; Reynaud, 2010). Although many ribonucleic acid (RNA) transcripts had been discovered throughout the years, they were only assumed to serve as intermediaries between DNA and proteins during gene expression. Within the last decade or so, many other components that are involved in different gene regulatory networks were unravelled and the passive role that was presumed to be played by RNAs is now considered to be misrepresentative in biology. Since the 1990s, different classes of RNA molecules were shown to be involved in gene expression and changes to the central dogma of molecular biology that was proposed by Crick now include our understanding of the roles that RNAs play in genome regulation today (Figure 1).

Several important discoveries of regulatory RNA molecules over the last decade revealed the diverse roles of RNA and triggered a gradual paradigm shift away from a protein-centric view in molecular biology (Higgs and Lehman, 2015; Prasanth and Spector, 2007). Beginning in the 1960s, scientists discovered that cell produce a large group of heterogenous nuclear RNAs (hnRNAs) that mainly reside in the cell nucleus and are degraded rapidly after synthesis. Only a small proportion of the RNA synthesized is exported out to the nucleus to interact with the ribosome complex in the cytoplasm (Harris and Watts 1962; Scherrer *et al.,* 1963; Soeiro *et al.,* 1966). It was hypothesized that these transcripts may be involved in gene regulation and could be interacting with the chromatin. However, little evidence was available due to technical limitations at that point in time (Britten and Davidson, 1969; Britten and Davidson, 1971). Transfer ribonucleic acids (tRNAs) were the first non-coding RNAs discovered in 1965 to have a function in protein synthesis. These molecules are capable of folding into structures and are involved in the protein translation process by base-pairing amino acids with the matching codon on the mRNA transcript (Holley *et al.,* 1965). The translation process takes place on the ribosome and when the tRNA adaptor encounters its corresponding codon, it transfers the amino acid that it was carrying to the end of the growing amino acid chain (Quigley and Rich, 1979). Today, advancement in tRNA research revealed the diverse roles of tRNAs beyond the process of translation. These highly versatile molecules are capable of adopting different post-transcriptionally modified profiles and have a role in maintaining cellular homeostasis (Dittmar *et al.,* 2006; Geslain and Pan, 2010; Parisien *et al.,* 2013).

The discovery of alternative splicing in the late 1970s greatly changed our perception that one gene led to the production of only one protein (Burge *et al.,* 1999; Caceres *et al.,* 1997; Morris and Mattick, 2014). Protein coding genes were found to be harbouring segments of coding (exons) and non-coding sequences (introns). During transcription, the pre-mRNA transcript copied from DNA contains both exonic and intronic regions. The non-coding regions undergo a process known as “splicing” and the adjacent exons are joined to form a contiguous coding sequence (Sharp, 2005). The splicing machinery is controlled by a group of RNAs known as small nuclear RNAs (snRNAs) that assembles in a ribonucleoprotein complex in the cytoplasm and are imported back into the nucleus to carry out their functions. The removal of introns and the joining of exons must be carried out precisely to prevent the formation of defective proteins (Wahl *et al.,* 2009; Will and Luhrmann, 2011). This landmark discovery earned Phillip Sharp and Richard Roberts the Nobel Prize for Medicine in 1993. Within the nucleolus, the production of mature and functional tRNAs, snRNAs, and

ribosomal RNAs (rRNAs) requires a different class of ncRNAs know as small nucleolar RNAs (snoRNAs). SnoRNAs have specific sequences and assemble into small nucleolar RNPs (snoRNPs) with a set of proteins in the cytoplasm (Dieci *et al.,* 2009). This snoRNP complex shuttles to the nucleolus to direct the modification of different groups of RNA molecules for them to be functional, such as in the case of the 2'-*O*-ribose methylation of rRNAs (Tollervey, 1997; Kiss, 2002; Züst *et al.,* 2011).

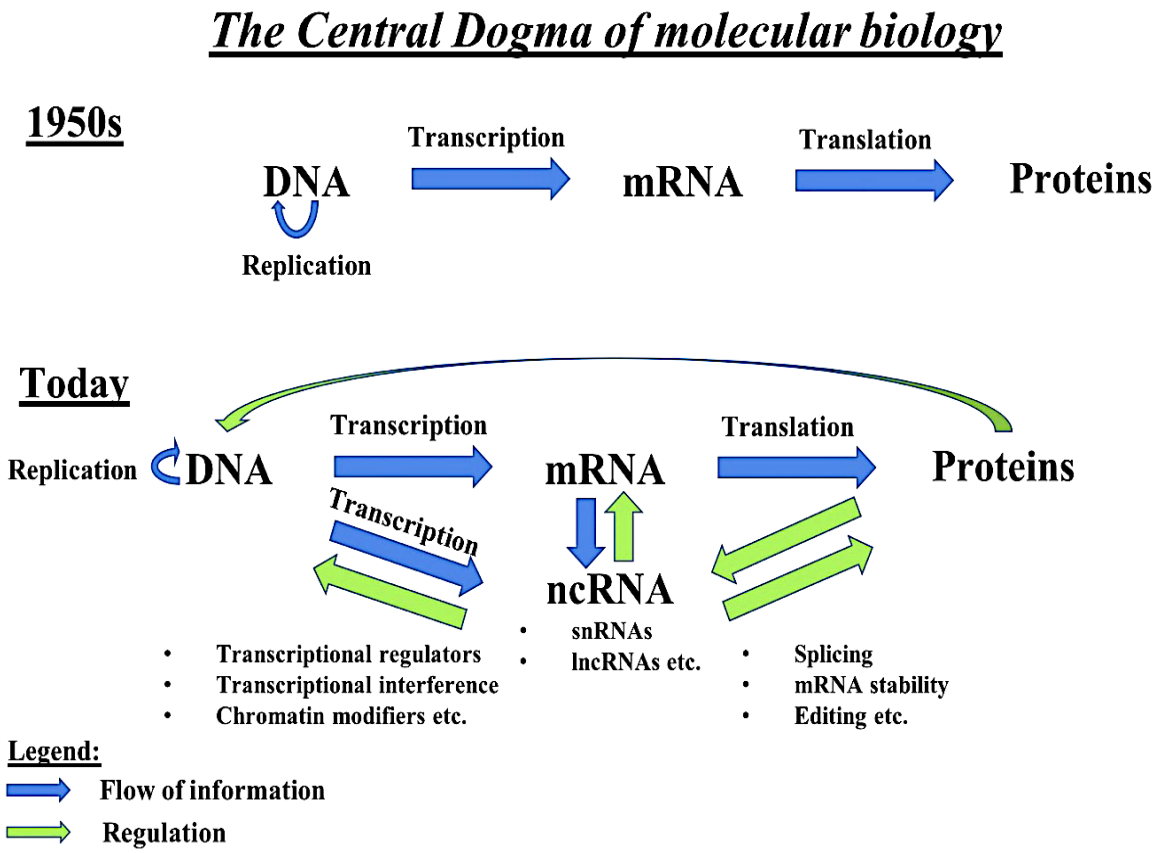


Figure 1. Gene Expression: From past to present.
The central dogma of gene expression was a concept introduced by Francis Crick to explain the route of DNA transcription and mRNA translation during protein synthesis. Extensive research in the last decades has revealed that the transcriptome is highly complex and pervasive. Many RNA transcripts do not code for proteins and instead they contribute to gene regulatory processes. These discoveries provided an important new perspective on the significance of RNA in gene regulation.

The discoveries of RNase P and riboswitches in the 1980s brought about a new wave of evidence that ncRNAs also have the ability to catalyse chemical reactions. Riboswitches are regulatory elements that form different secondary structures upon ligand binding and this allows them to regulate gene expression without the need for any proteins (Nudler and Mironov 2004; Tucker and Breaker 2005; Winkler, 2005). They are important RNA thermosensors that are activated in response to environmental or metabolic signals (Johansson *et al.,* 2002). A Nobel Prize in Chemistry was awarded to Sydney Altman and Thomas Cech

for their breakthrough in the catalytic properties of RNAs. Today, emerging evidence from different organisms is revealing that many RNA transcripts not only relay important informational genetic material in the form of messenger RNAs but also serve in catalytic and regulatory capacities (Soshnev *et al.,* 2011; Briggs *et al.,* 2015; Spadaro *et al.,* 2015). The completion of the human genome project (HGP) and subsequent annotation efforts indicated that approximately 1% of the human genome comprises of protein-coding sequence, whereas the rest of the genomic sequences were widely regarded as "junk" or non-coding DNA (Lander *et al.,* 2001; Venter *et al.,* 2001; Ponting *et al.,* 2009). Although it is currently widely acknowledged that RNAs are involved in various biological processes, the total number of ncRNAs, classifications and functions of many of these transcripts are still highly debated even today.

1.2 Non-coding RNAs: multi-functional cellular molecules

The human genome can be considered to be more than just the transcription of protein-coding sequences and the production of some non-coding RNAs (e.g., tRNA, rRNA, and snoRNA) that have important biological functions in this post-genomic era. One of the most important discoveries since the publication of the DNA structure was the HGP that provided us with a glimpse of the genetic blueprint that makes up a human (Lander *et al.,* 2001; Venter *et al.,* 2001). What came as a surprise in this report for many was the diversity of transcriptional products that are produced from the genome. A variety of transcriptomic studies were performed since then and it was indicated that expression of the mammalian genome is highly pervasive with transcription from both strands of the DNA and initiation of transcription from different transcriptional start sites (TSS) (Hayashi *et al.,* 2001; Okazaki *et al.,* 2002; Carninci *et al.,* 2005; Mattick and Makunin, 2006; Djebali *et al.,* 2012; Zhao *et al.,* 2016). Transcripts were found to be mapping to intronic, intergenic and antisense regions which added another layer of complexity to the transcriptome (Okazaki *et al.,* 2002; Katayama *et al.,* 2005).

Data from the ENCODE project estimated that approximately >80 % of the human genome was actively transcribed with a majority of these transcripts produced lacking protein-coding capacity and were thus named 'non-coding RNAs' (ncRNAs) (The ENCODE Project Consortium, 2012). Even protein-coding genes are mainly composed of intronic sequences and only 1 % of the genome is considered to be coding for exons (Tripathi *et al.,* 2010, Tripathi *et al.,* 2013). The recognition that vast numbers of ncRNA transcripts are

generated by the genome resulted in countless debates about their roles in development and evolution. Irrespective of the developmental complexity of the organism, the number of protein-coding genes has been shown to be constant across vertebrates and metazoans. On the other hand, the number of ncRNAs was increasing with developmental complexity and it was proposed that they had crucial roles in regulatory processes and perhaps related to the complexity of an organism (Figure 2) (Ulitsky *et al.,* 2011; Amaral *et al.,* 2008; Mattick, 2004).

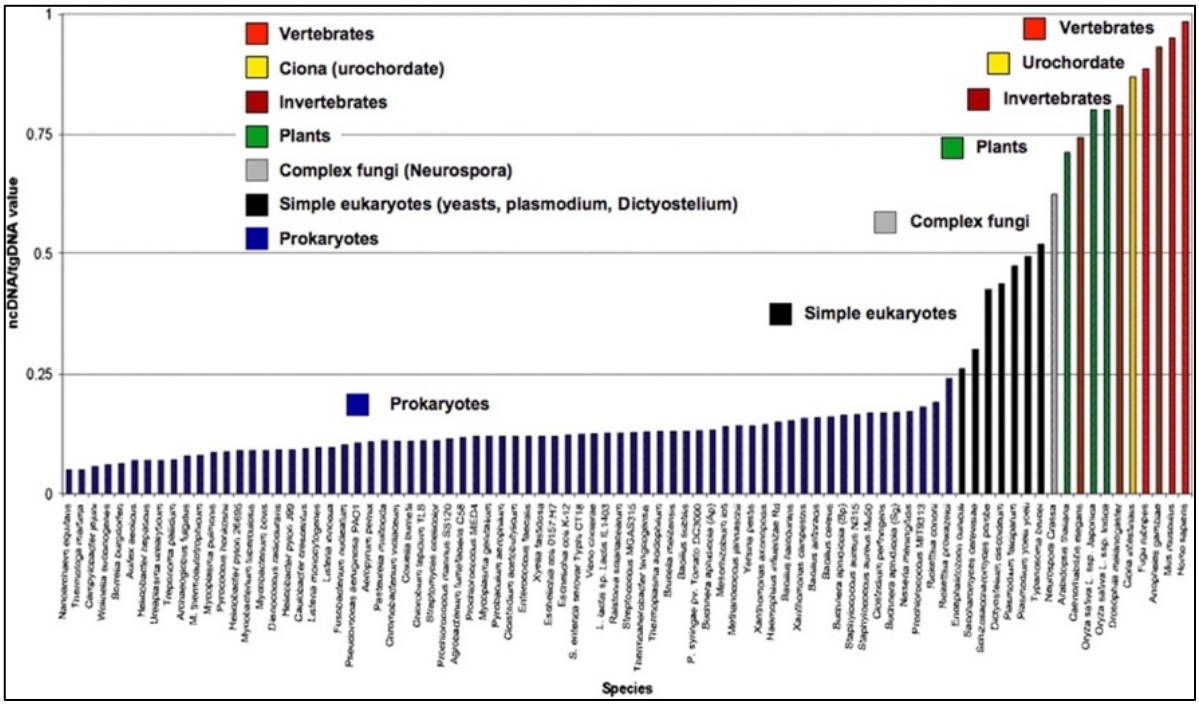


Figure 2: The proportion of ncRNAs expands with developmental complexity.
There is a wide range in genome sizes from the simplest to the most complex organism. The complexity of an organism increases with the amount of non-coding RNAs in the genome rather than protein-coding genes. [Image taken from Taft *et al.,* 2007].

Advances in sequencing technologies such as high-throughput sequencing brought to light the existence of many different classes of ncRNAs that serve vastly different roles within the cell (Figure 3). During development and reproduction, the normal function of cells is dependent on precise expression of both proteins and non-coding RNAs. Precise control of gene expression is essential for all organisms. Examples of gene regulation events include transcription, mRNA splicing, mRNA export, mRNA stability, translational control, and RNAi (Fire *et al.,* 1998). Two major classes of ncRNAs have been shown to be important in different biological processes. Here, they are distinguished based on the length of the mature transcript, namely small RNAs (sRNAs, which are shorter than 200-nt) and long non-coding RNAs (lncRNA, larger than 200-nt in size). This form of classification was used to distinguish the sRNAs from the longer non-coding transcripts and can be considered to be

arbitrary. Extensive studies carried out in the last decade had shown that sRNAs were crucial in a plethora of biological process including the regulation of gene expression as well as transcriptional and post-transcriptional gene silencing (Pauli *et al.,* 2011).

Well-studied sRNAs include microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs), and PIWI-interacting RNAs (piRNAs). As the core component of the RNAi pathway, these sRNAs repressed the expression of protein-coding genes. Single-stranded sRNAs, typically 19- to 30- nt in length, form a complex with proteins of the Argonaute family. The specificity of small RNA pathways is determined by the small regulatory RNA, which guides the Argonaute proteins via complementary base pairing to its target RNA. These sRNA pathways are able to regulate gene expression at the transcriptional or post-transcriptional level, as well as via translational repression (Bartel, 2004; Lippman and Martienssen, 2004; Chapman and Carrington, 2007; Wu and Belasco, 2008). RNAi was shown in many studies to mediate resistance to parasitic nucleic acids such as transposons and viruses (Hannon, 2002) (Figure 3).

Lin-4 and *let-7* were the first miRNAs to be identified in *Caenorhabditis elegans.* miRNAs are the most extensively studied class of ncRNAs and since their discovery, hundreds of miRNAs have been identified to be involved in various developmental processes including control of cell proliferation, apoptosis, differentiation and development (Bartel, 2004; Stefani and Slack; 2008). Changes in the sequence or expression levels of miRNAs have been shown to play a part in many human diseases such as cancer (Iorio *et al.,* 2005; Yanaihara *et al.,* 2006). The field of miRNAs expanded rapidly with thousands of transcripts annotated in 271 species on miRbase (version 22), including humans (Kozomara and Griffiths-Jones, 2011, Kozomara *et al.,* 2019). In the case of endogenous siRNAs, they were first reported from studies in plants, fungi and *Caenorhabditis elegans* (Hamilton and Baulcombe, 1999; Baulcombe, 2003; Pak and Fire, 2007). A mature dsRNA (~22-nt in length) was formed after processing by the RNAse III enzyme Dicer (Ketting *et al.,* 2001; Bernstein *et al.,* 2003; Murchison *et al.,* 2007). One of the strand of the dsRNA (guide strand) associates with the RNA-induced silencing complex and it will bring the complex to a complementary target RNA to direct its cleavage. (Song *et al.,* 2004; Rivas *et al*., 2005). Fire and Mello were awarded a Nobel Prize for their discovery of short RNA interference in 2006. The third class of sRNAs are piRNAs, which were first observed in *Drosophila melanogaster* for transposon silencing (Aravin *et al.,* 2007; Brennecke *et al.,* 2008; Czech *et al.,* 2013; Czech *et al.,* 2018). Most of the sequences that belongs to this class of sRNA (24- to 29-nt)

were found to map to transposons and genomic repeats and they interact with Piwi-clade Argonaute proteins to maintain genome stability in germline cells (Figure 4) (Saito *et al.,* 2006; Vagin *et al.,* 2006; Brennecke *et al.,* 2008; Gunawardane *et al.,* 2007). Research in the piRNA field has grown progressively over the years with some studies also in stem cell and cancer biology (Gonzalez *et al.,* 2015; Cantarella *et al.,* 2012; Fagegaltier *et al.,* 2016).

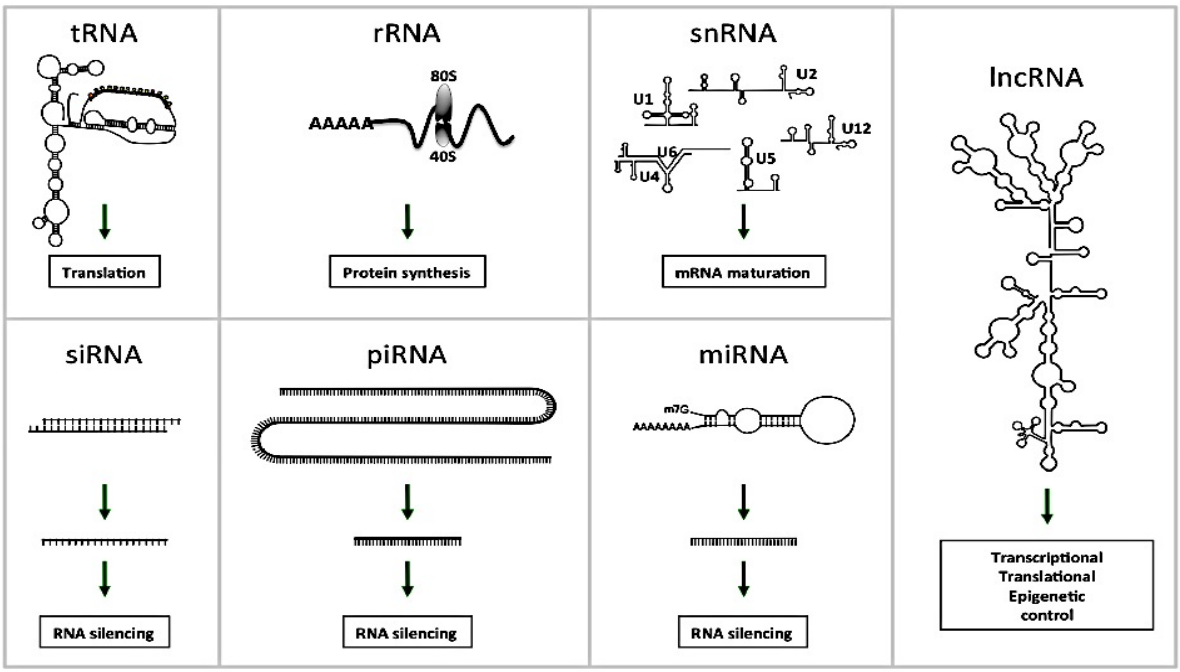


Figure 3: Schematics of described animal RNA transcripts without protein-coding potential. These include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), long non-coding RNAs (lncRNAs) as wells as variety of small non-coding RNAs such as endogenous small interfering RNAs (endo-siRNAs), PIWI-interacting RNAs (piRNAs) and microRNAs (miRNAs). [Image taken from Morceau *et al.,* 2013].

Besides sRNAs, another subset of ncRNA transcripts were shown to have important regulatory functions in the genome. Large transcripts known as lncRNAs were reported even before the identification of sRNAs with the discovery of *H19* in 1988 and *Xist* in 1991 (Pachnis *et al.,* 1988, Borsani *et al.,* 1991). LncRNA *H19* was shown to be a highly abundant lncRNA and was first indicated to be involved in muscle development. This imprinted lncRNA gene was found to be expressed from the maternal allele and regulates the expression of insulin-like growth factor 2 (IGF2) (Nordin *et al.,* 2014; Wilkin *et al.,* 2000). Following the discovery of lncRNA *H19*, another lncRNA was found to function as an epigenetic switch during X chromosome inactivation. (Brown *et al.,* 1992; Ballabio and Willard, 1992). The X inactive specific transcript (*Xist*) silences one of the X-chromosomes and renders it inactive by spreading itself along the X chromosome and recruiting the Polycomb repressive

complexes 1 and 2 (PRC1/PRC2) along the inactive X chromosome (Borsani *et al.,* 1991; Brown *et al.,* 1992; Brockdorff *et al.,* 1992).

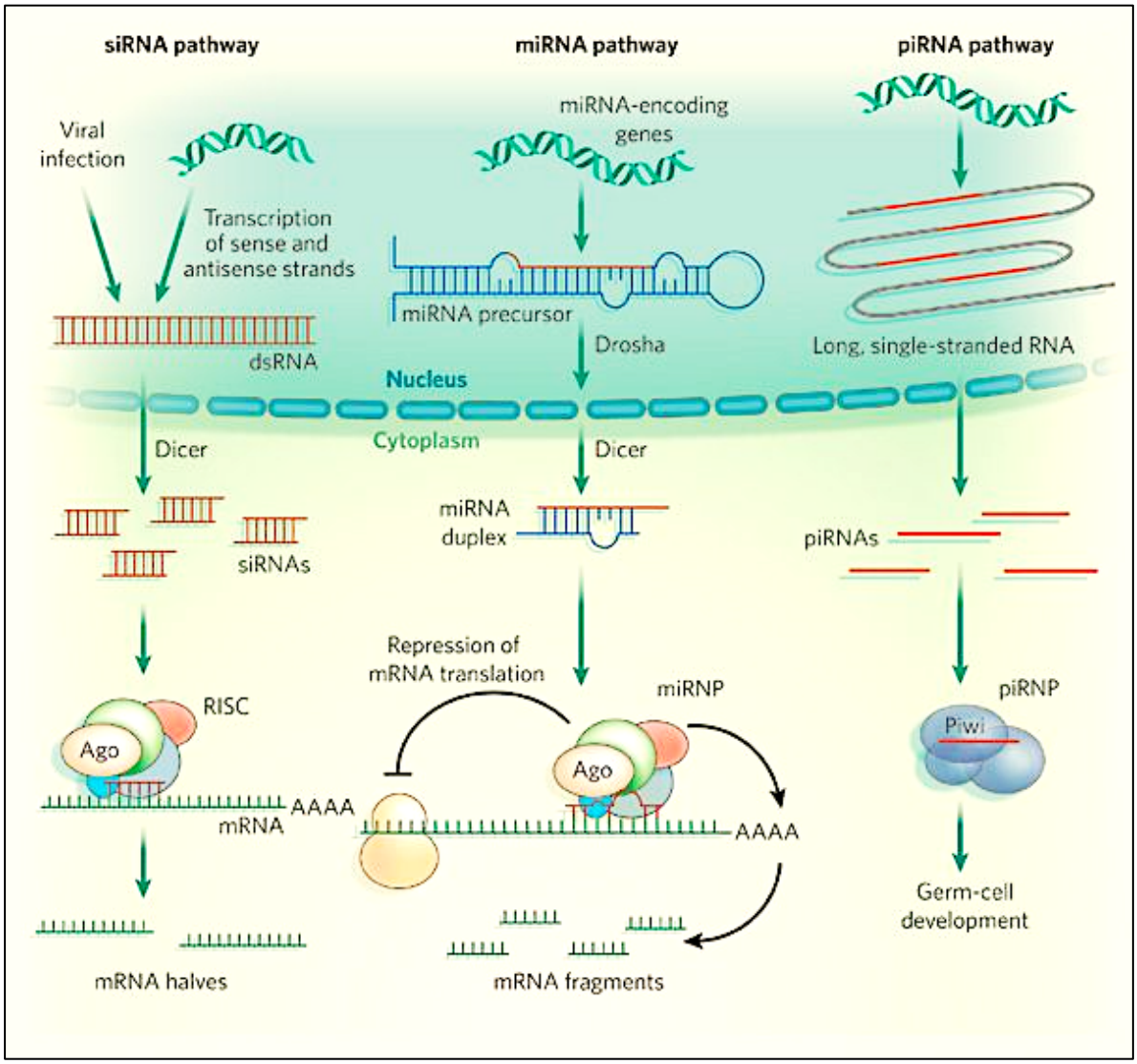


Figure 4: Basic outline of three different regulatory sRNAs pathways.
MicroRNAs, endo-siRNA and piRNAs undergo different biogenesis pathways and differ in their substrates recognition and target regulation. An siRNA duplex containing a guide strand and a passenger strand is form after Dicer-2 processes the dsRNA precursor. The guide strand is loaded onto Argonaute2 (AGO2) protein and guide the protein to cleave target RNA transcripts. Mature miRNA strands are formed from the processing of primary miRNAs (pri-miRNA) and precursor miRNA (pre-miRNA) by Dicer-1 and Argonaute1 (AGO1) protein respectively. The miRNA-AGO1 complex binds to RNA transcripts and regulates gene expression post-transcriptionally. piRNAs form the third major class of small RNAs and one of their function is to silence transposons within the germline. [Image taken from Grosshans and Filipowicz, 2013].

The first insight of lncRNAs that were involved in epigenetic phenomena had prompted researchers to search for other functional lncRNAs that were involved in developmentally regulated processes. Since then, the extent of lncRNA-related research has risen with studies in different animal models and diseases (Figure 5). Other examples of well-studied lncRNAs that have tissue-specific and developmental expression patterns includes

Tsix, MALAT1, HOTAIR, ANRIL, NEAT1 and Braveheart (Lee *et al.,* 1999; Ji *et al.,* 2003; Rinn *et al.,* 2007; Broadbent *et al.,* 2008, Sunwoo *et al.,* 2009; Klattenhoff *et al.,* 2013). In summary, the discovery of ncRNAs has underlined the surprising roles and functions of RNA-based regulatory control and offered fundamental insights into novel mechanisms that were involved.

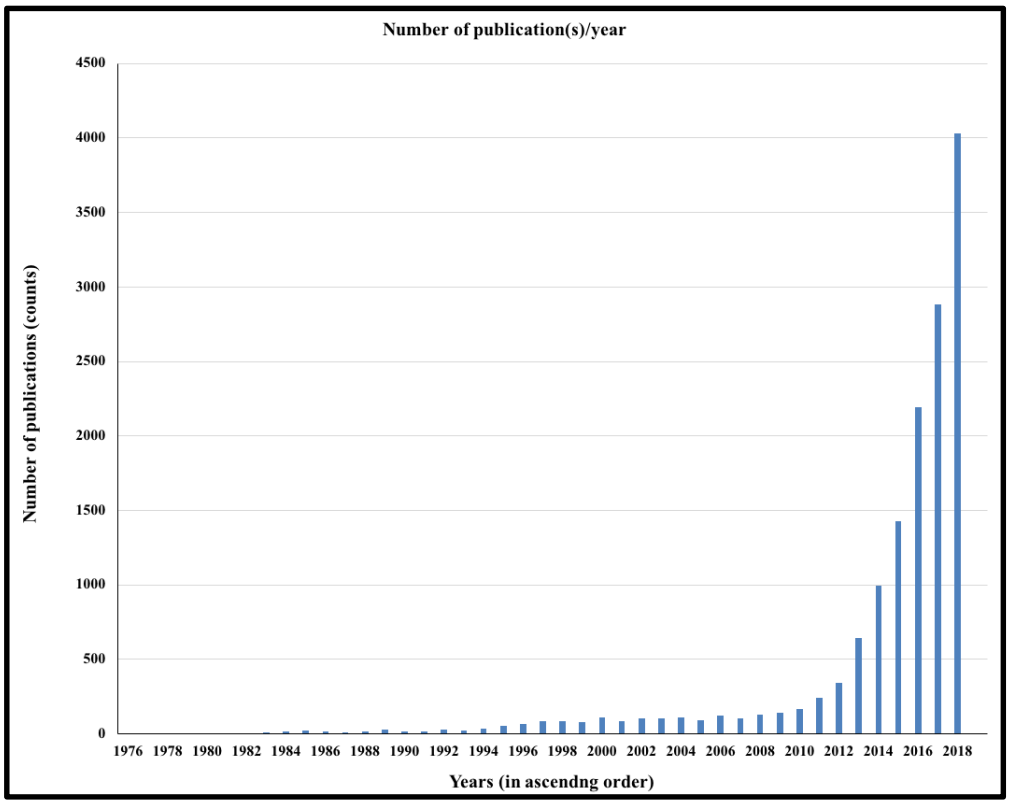


Figure 5: Number of lncRNA-related publications since the late 1970s.
The number of queried publications is obtained from PubMed library with keyword 'lncRNA' or 'long non-coding RNA'. Since the discovery of lncRNA *H19* and *Xist*, there is a surge in the number of publications in this field. This upward trend reflects the increase in interest in this area of research and is made possible with the advancement of biochemical and molecular technologies.
[Data accessed on 30th January 2019].

1.3 Identification and Characterisation of lncRNAs

The introduction of high-throughput sequencing methods has revolutionized and accelerated the identification and annotation of ncRNAs in different model organisms (Akay *et al.,* 2019, Wen *et al.,* 2016; Zhao *et al.,* 2018). In order to discover the number of lncRNA transcripts in the genome, a combination of different biochemical and molecular biology methods have been used to map the mammalian transcriptome. An international effort was established for the first time to generate the first lncRNA collection from the mouse and human transcriptome (Carninci *et al.,* 2005; Derrien *et al.,* 2012). The FANTOM (Functional Annotation of The Mouse) consortium and GENCODE project (that is part of the ENCODE -

ENCyclopedia Of DNA Elements - in humans) performed deep-sequencing of large pools of cDNA clones from mouse and human and reported the identification of thousands of lncRNA transcripts. The number of protein-coding genes generated from the genome across different organisms remains relatively constant but the number of lncRNA transcripts discovered continues to increase. Statistics first release by the GENCODE v7 consortium identified 14,880 human lncRNAs in 2012. In the latest release of GENCODE v29, a total of 58,721 genes were transcribed within the human genome, of which ~28 % (16, 066) were labelled as lncRNA transcripts.

Besides the improvement in sequencing methods, the development of innovative methods and bioinformatics pipelines provided scientists with better tools to assemble the transcriptome and enhanced the process of identifying genes and annotating them in the genome. Technologies to map TSSs such as CAGE (cap analysis of gene expression) and RAMPAGE (mapping of promoters for analysis of gene expression) and the development of RNA-sequencing technologies not only allowed us to identify novel transcripts but also construct their full-length sequence and identify the exon/intron boundaries of the transcripts (Guttman *et al.,* 2009, Sun and Kraus, 2015, Trapnell *et al.,* 2010). Today, lncRNAs have been identified across many organisms such as humans, mice, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Dario rerio* (Cabili *et al.,* 2011; Guttman *et al.,* 2009; Brown *et al.,* 2014; Yamashita *et al.,* 2016; Nam and Bartel, 2012; Swiezewski *et al.,* 2009; Kaushik *et al.,* 2013). The collection of data available through various systematic databases such as lncRNAtor (Park *et al.,* 2016), GENCODE (Derrien *et al.,* 2012), LNCipedia (Volders *et al.,* 2013), NONCODE (Zhao *et al.,* 2016) and lncRNAdb (Amaral *et al.,* 2011) served as a rich resource for the acceleration of lncRNA-related research.

The definition of lncRNAs remains fluid as compared to sRNAs and it is constantly evolving with new insights being made in the field. LncRNA molecules were shown to share many features with messenger RNAs (mRNAs) except that they are non-coding transcripts (Figure 6). They are transcribed by RNA polymerase II and most of the transcripts undergo processing such as 5′-capping, splicing, and polyadenylation (Gomes *et al.,* 2013). They are usually defined as transcripts with length greater than 200-nt that lack significant protein coding potential (Rinn and Chang, 2012, Sun and Kraus, 2013). These transcripts can be generated from intergenic regions, promoter regions and sequences overlapping or in an antisense orientation to annotated protein-coding genes (Gomes *et al.,* 2013). Instead of being

used as templates for protein synthesis, lncRNAs are able to carry out various biological functions. In contrast to small ncRNAs, which typically feature processed transcripts of defined length, long non-coding RNAs vary dramatically in their size, with RNAs up to several mega-bases in length (Dinger *et al.,* 2008). Although the expression of lncRNAs is regulated spatiotemporally, their expression levels are typically significantly lower compared to those of protein-coding mRNAs (Tsoi *et al.,* 2015).

Although these features had been widely applied in the field to define lncRNAs, this has led to many false positive and false negative results. For example, it has been shown that lncRNAs consist of both polyadenylated and non-polyadenylated transcripts (Hangauer *et al.,* 2013). Similarly, studies have suggested that many lncRNAs harbour the capacity to encode miniORFs that could be functional (Andrews and Rothnagel, 2014; Bazzini *et al.,* 2014). Examples of lncRNAs that encode for small peptides and could have been placed in the protein-coding gene classification include lncRNA *Xist*, lncRNA *ENOD40* and lncRNA *HOXB-AS3* (Dinger *et al.,* 2008; Matsumoto *et al.,* 2017; Huang *et al.,* 2017). Several computational methods such as Coding Potential Assessment Tool (CPAT) and Phylogenetic Codon Substitution Frequency (PhyloCSF) were generated to tackle this issue and they were commonly used as a filtering step to identify putative lncRNA transcripts. These methods rely heavily on the properties on known lncRNAs as training datasets and could be biased during selection. With the increasing amount of sequencing data generated, a more specific and sensitive computational algorithm is required to separate coding genes from non-coding transcripts. Another option to define lncRNAs and distinguish them from miniORFs and small peptides is ribosome profiling or shotgun proteomics (Ingolia *et al.,* 2012; Giambruno et *al.,* 2018)

Evolutionary conservation has always been used to understand the functions of protein-coding genes. In comparison to protein-coding genes, many lncRNAs lack conservation of their primary sequences. However, this criterion should not be mistaken for a lack of function in lncRNAs (Diederichs, 2014). LncRNA *Xist* and *Megamind/Tuna* are examples of lncRNAs that were highly conserved in terms of function across zebrafish, mouse and human yet displayed low sequence conservation across the species (Brockdorff *et al.,* 1992, Utilisky *et al.,* 2011, Lin *et al.,* 2014). LncRNA *Megamind/Tuna* was an ultra-conserved lncRNA that was first identified in zebrafish that had important neural development function. This observation suggested that sequence conservation may not be as an important factor for RNA function compared to proteins. Furthermore, a comparison of

lncRNAs in different species has revealed that lncRNAs stem from syntenic loci between different species. The order of syntenic transcripts along the chromosomes in different species is conserved (Herriges *et al*, 2014).

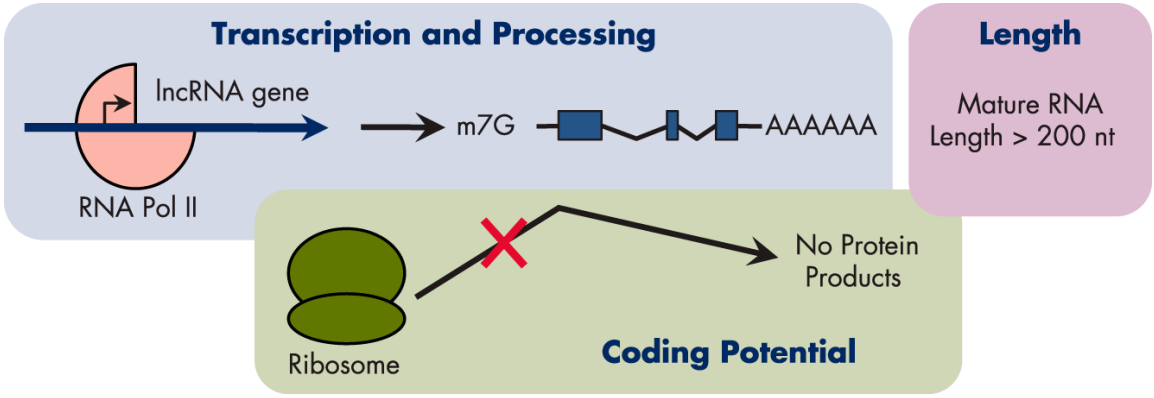


Figure 6: Properties of long non-coding RNA
The mammalian genome is pervasively transcribed by RNA Polymerase II, generating both protein-coding messenger RNAs (mRNAs) and long non-coding RNAs transcripts of more than 200 nucleotides. These strands of lncRNA are mostly poly-adenylated and capped at the 5' end. The majority of these ncRNA transcripts also lack conserved or long open reading frames (ORFs). [Image taken from Sun and Kraus, 2014].

In conclusion, the task of distinguishing lncRNA transcripts within the transcriptome could is complex and challenging as the current definition may not be the best to define a lncRNA transcript. A good example to illustrate this concept is the lncRNA steroid receptor RNA activator (*SRA1*). The human *SRA1* was first identified as a lncRNA that is involved in transcriptional regulation and associated with human diseases such cancers (Lanz *et al.,* 1999; Cooper *et al.,* 2009; Liu *et al.,* 2016; Yan *et al.,* 2016; Kyung *et al.,* 2017). Interestingly, this bifunctional locus that codes for the lncRNA *SRA* also encodes for the *SRA* protein (SRAP). Both *SRA* and *SRAP* have been shown to be involved in interacting with a variety of proteins, demonstrating the presence of bifunctional RNAs that have the ability to exert coding- and coding-independent functions (Nam *et al.,* 2016; Liang *et al.,* 2018). For these reasons, the identification of lncRNAs should to be accompanied by a range of experiments in order to annotate them accurately.

1.4 Classification of lncRNAs

1.4.1 Cataloguing lncRNAs based on genomic location

With the increasing number of lncRNAs being discovered and thought to be involved in a myriad of biological processes, it is vital to organise them into categories to have a better

understanding of their biological significance. One common way to distinguish the various types of lncRNA transcripts is to categorize them based on their genomic origin. Due to our poor understanding of this class of ncRNA, thousands of RNA transcripts fall into the class of lncRNAs due to the size selection criteria for lncRNA transcripts. They are annotated as sense, antisense (NAT), bidirectional, intronic, intergenic lncRNAs (Khandelwal *et al.,* 2015). This form of annotation considers the position of the transcript relative to annotated protein-coding genes (Figure 7). However, it is not mutually exclusive and some transcripts can be classified into two or more categories. An example would be MALAT1, an intergenic, antisense lncRNA transcript (Louro *et al.,* 2009).

Sense transcripts are transcribed on the same strand as protein-coding genes whereas natural antisense transcripts (NATs) are located on the opposite strand. The sense transcripts can be further divided into sense overlapping or sense intronic lncRNAs. An example of a sense lncRNA would be *SNHG4*. NATs are considered to be one of the most abundant forms of lncRNA transcripts (Kung, 2013). These lncRNAs are endogenous molecules that partially overlap with their sense counterparts. There are many examples of NATs in the genome and some well-known examples include *BACE1-AS* and *MALAT1* that will be discuss in the later chapters. On the other hand, bidirectional refers to transcripts that are expressed on the opposite strand of an annotated protein-coding transcript. The transcription start site of these transcripts is usually <1,000-nt away from protein coding genes (Liang *et al.,* 2014). Although the term lncRNA and long intervening RNA (lincRNA) is often used interchangeably, it should be noted that the latter is a subclass of lncRNAs derived from regions between known protein-coding genes. In order to qualify as a lincRNA, transcripts should be outside of known protein-coding genes, harbour chromatin methylation marks such as H3K4me3 and H3K36me3 and contain <35 % open reading frame (ORFs) in the sequence (Khalil *et al.,* 2009). Another class are intronic lncRNAs. These transcripts can be in sense or antisense orientation and they are embedded within intronic regions of protein-coding genes (Nakaya *et al.,* 2007).

Several studies have shown that intronic lncRNAs have autonomous transcriptional regulation and they are expressed independently from the protein-coding genes that they reside in (Louro *et al.,* 2009; Laurent *et al.,* 2015). However, lncRNA-related research is a relatively new field and many of these transcripts do not just fit into one particular category. With the vast number of genome sequencing projects that have been carried out, it is now apparent that the genome produces a large number of transcripts, both coding and non-coding,

with a multitude of isoforms. Similarly, lncRNAs that are transcribed have low expression levels and undergo the process of alternative splicing and thus making it challenging to assemble the full-length transcripts. A better understanding of the transcriptome is required in order to better annotate lncRNAs as a functional class of molecules.

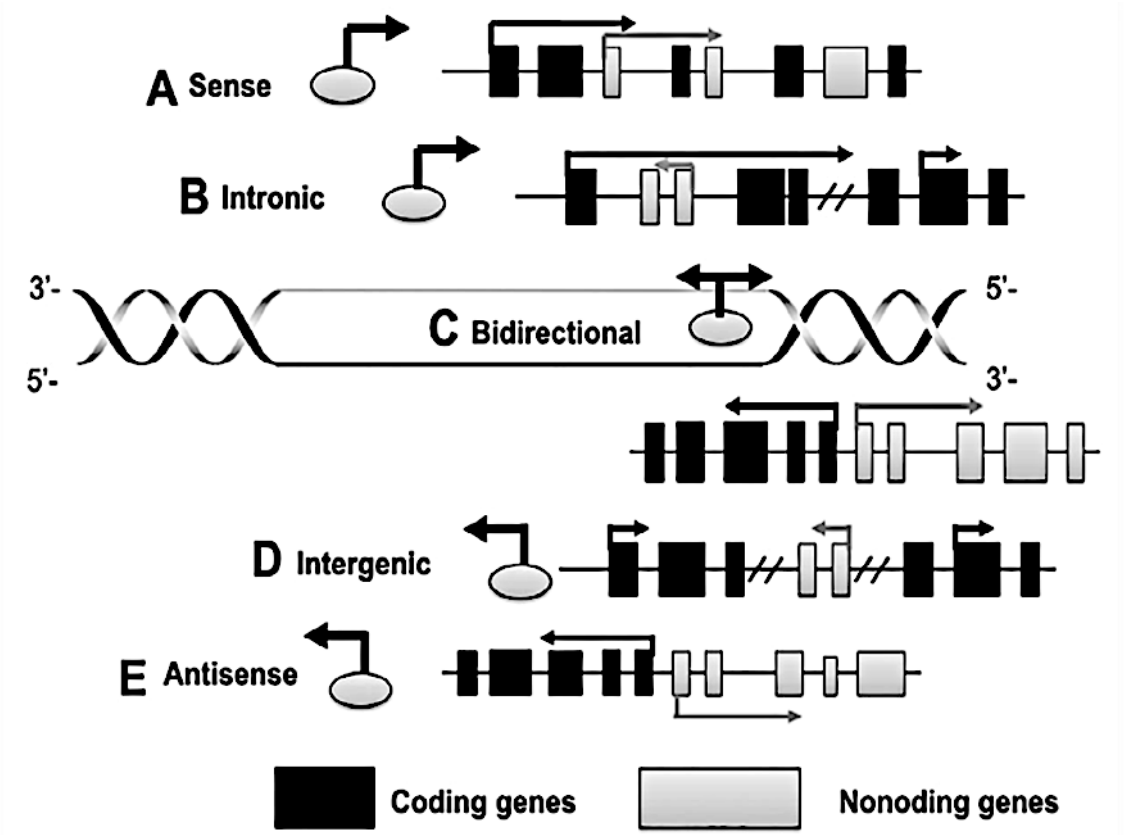


Figure 7: A broad overview of different forms of lncRNA transcripts.
LncRNA transcripts can be annotated as sense, antisense, bi-directional, intronic or intergenic. (A, B) With reference to the location of protein genes, sense and antisense lncRNA transcripts are usually found to be overlapping with the exons of protein-coding genes or within the exonic regions. A sense lncRNA is transcribed from the same strand as the protein-coding gene, whereas an antisense one from the opposite strand. (C) A ncRNA is known as a bidirectional lncRNA when its TSS is on the opposite strand and is located within 1-kb from the TSS of a protein-coding gene. (D) Intronic lncRNAs refers to transcripts that are found within the intronic regions of protein-coding genes. (E) LncRNA transcripts that are transcribed from genomic regions between separate gene loci are classified as intergenic lncRNAs, or sometimes also known as intervening lncRNAs (lincRNAs). [Image taken from Liu *et al.,* 2018]

1.4.2 Classification based on molecular mechanisms of lncRNAs

The discovery of ncRNAs and their mode of actions expanded our understanding of gene regulation. Compared to lncRNAs, the molecular modes of action of sRNAs are well established. For many years, the research community has been divided on whether lncRNAs are functional or just represent transcriptional noise (Blake *et al.,* 2003). Recent work in a few

of the lncRNAs have demonstrated that these transcripts are important mediators of cellular functions acting at different levels (Pintacuda *et al.,* 2017). Given the flexibility of lncRNA strands, it has been proposed that lncRNA transcripts are able to interact with DNA, RNA and protein partners to form DNA: RNA hybrids; RNA: RNA hybrids and RNA-protein interactions so as to modulate different biological processes (Rinn and Chang, 2012). In recent years, the formation of RNA-DNA triplex has been proposed to be the fourth type of interaction between an RNA transcript and the double-stranded DNA (Yue *et al.,* 2016). Through the different types of interaction between a lncRNA and its partners, a wide range of regulation is carried out and they can be divided into four classes – molecular decoys, molecular scaffolds, molecular signals and molecular guides. Given the versatility of lncRNAs, they are not restricted to a fixed mode of action and some have been shown to have more than one mechanism of action (Wong and Chang, 2011; Shakil *et al.,* 2016) (Figure 8).

Molecular decoys

Regulation of transcription is essential for proper development of an organism. During transcription, lncRNAs can influence gene expression either positively or negatively. LncRNAs that function as a decoy behave like a molecular "sink" and hinders the action of transcriptional regulators such as transcription factors (TF) or post-transcriptional regulators such as miRNAs (Figure 8a). One well-studied example is the inhibition of the NF-YA nuclear transcription factor activity by lncRNA *PANDA* (Hung *et al.,* 2011). The function of *PANDA* is to impede the expression of apoptotic genes during DNA damage, thereby preventing activation of apoptosis and promoting cell survival. Its interacting partner, NF-YA, is a nuclear transcription factor that normally drives the expression of apoptotic genes. During cell cycle arrest, *PANDA* acts as a decoy and prevents NF-YA from binding to apoptotic genes and thereby preventing apoptosis (Hung *et al.,* 2011). Besides proteins, lncRNAs can also sequester miRNAs and prevent them from carrying out gene silencing (Poliseno *et al.,* 2010). The 3' UTR of lncRNA *PTENP1* possess similar binding sequences for regulatory miRNAs that targets PTEN, a tumour suppressor gene. Binding of lncRNA *PTENP1* to regulatory miRNAs sequesters their availability to PTEN mRNA, which in turn increases the availability of PTEN protein in the cell (Poliseno *et al.,* 2010; Song *et al.,* 2011).

Molecular signals

The second mode of action of lncRNAs is through cellular signalling. Cells have the ability to process different stimuli and perform a series of changes to the cellular state through a wide range of different signalling pathways (Figure 8b). LncRNAs in this category are important biomarkers to coordinate spatio-temporal events during various biological processes (Peng *et al.,*2017). An example of a signal lncRNA is lincRNA-p21. In the event of DNA damage, lincRNA-p21 functions as a signalling molecule and triggers apoptosis (Tran *et al.,* 2015). LincRNA-p21 is located upstream of the *CDKN1A* gene and its activity is tightly regulated by p53 (Huarte *et al.,* 2010). The above mentioned lncRNA *PANDA* does not just function as a molecular decoy, but it is also able to act as a signalling molecule during DNA damage. The induction of DNA damage leads to the expression of a set of signalling lncRNAs and one of them is the lncRNA *PANDA*. Its activity is p53-dependent and requires p53 activation during DNA damage (Hung *et al.,* 2011; Baldassarre and Masotti, 2012). Both lncRNA *PANDA and* lincRNA-p21 regulate and fine tune the process of apoptosis by keeping the anti-apoptotic and pro-apoptotic activity in check (Hung *et al.,* 2015; Dahl et al., 2018).

Molecular guides

The third mode of action for lncRNAs is molecular guides where they interact with specific proteins and guide them to specific DNA targets (Moran *et al.,* 2012; Froberg *et al.,* 2013). Guide lncRNAs are essential to help ribonucleoprotein (RNP) targeting to specific locations. The resulting lncRNA-protein complex is able to influence gene expression either in cis or in trans. Target proteins recognized by the lncRNA may be a repressor or activator complex, consequently transcription can be influenced negatively or positively, respectively. Transcriptional repression is one of the best understood functions of lncRNAs. For example, *Air* is a cis-acting lncRNA that interacts with the promoter of target genes on the paternal chromosome (Nagano *et al.,* 2008; Braidotti *et al.,* 2010). The accumulation of lncRNA *Air* on its target locus results in the recruitment of G9a methyltransferase. This enzyme then methylates histone H3 on lysine 9 (H3K9), which leads to transcriptional silencing (Nagano *et al.,* 2008; Mozzetta *et al.,* 2014).

Molecular scaffolds for other molecules

The last category comprises of lncRNAs that function as an anchoring structure for regulatory proteins. Some lncRNA transcripts are able to form secondary and tertiary structures and these structures contain multiple domains that allow interaction with protein partners (Yoon *et al.,* 2013). These RNP complexes bind to target loci and regulate gene expression. For example, the HOX transcript antisense RNA (*HOTAIR*) interacts with proteins on both the 5' and 3' end. The first 300-nt on the 5' end of the transcript are able to interact with Polycomb Repressive Complex 2 (PRC2) and inhibit transcription by methylation of H3K27 (Tsai *et al.,* 2010). On the other hand, the 3' end of *HOTAIR* activates gene expression when bound to LSD1/CoREST/REST protein complex (Tsai *et al.,* 2010). This phenomenon suggests that lncRNAs are able to act as a scaffold for multiple protein complexes and can result in the deposition of different histone modifications (Figure 8d).

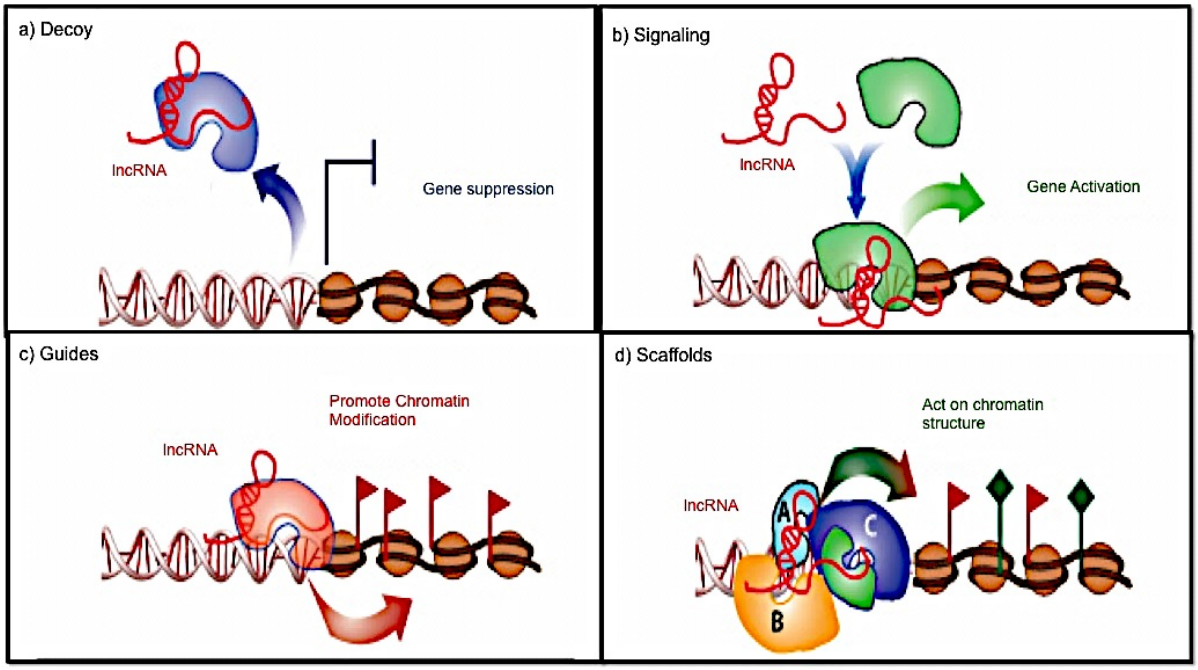


Figure 8: Schematic diagram depicting four modes of action by lncRNAs.
a) Molecular decoy: lncRNAs function as a molecular sponge that tethers RNA/protein factors and prevents these factors from binding to target genes. b) Molecular signal: Upon receiving an external stimulus, lncRNAs are used as biomarkers to rely in signalling pathways to activate gene expression. c) Molecular guides: lncRNAs that function as guides recruit chromatin modifying complexes to target genes in cis or in trans. d) Molecular scaffolds: lncRNAs act as a platform for multiple proteins to form a RNP complex for downstream processes such as histone modification.
[Image modified from Wang *et al.,* 2011].

1.4.3 From actions to functions - The diverse roles of LncRNAs

LncRNA research is an exciting and fast-moving area in the non-coding RNA world with new discoveries and annotations of lncRNA transcripts in different biological systems. One of the challenges of detecting and deciphering functional lncRNAs is due to their low expression levels and restricted spatio-temporal expression. Certain lncRNAs are only expressed in specific cell types, tissues or conditions such as stress or specific developmental time points. Therefore, understanding the cellular localisation of each lncRNA transcript will provide a clue into their functions. With the improvements in biochemical and molecular tools in the last few years, we now have a better understanding of this new class of molecular regulators and their molecular mechanisms of action and biological functions within the cell. LncRNAs can be enriched in either the nucleus, cytoplasm or present in both, and depending on the compartment in which they are located, they play diverse roles ranging from cis- to trans-regulators of gene expression in the nucleus to post-transcriptional regulators in the cytoplasm (Miyagawa *et al*., 2012; Lubelsky and Ulitsky, 2018; Shukla *et al*., 2018). I will discuss some of the roles of lncRNAs and it should be noted that these are just the tip of the iceberg and focused mainly on the RNAs as the functional units themselves. As the field continues to mature, I expect the categories to expand to encompass the diverse roles that lncRNAs play.

Epigenetic Regulators and chromatin organisers

Since the discovery of lncRNAs, their capacity to facilitate epigenetic regulation has been a hallmark of lncRNA research. Many studies have shown that lncRNAs are central in regulating allelic expression and are implicated in processes such as dosage compensation and imprinting (Davidovich and Cech, 2015). Amongst the many processes lncRNAs are involved in, their association with proteins of the polycomb group repressive complex (*PRC*) and trithorax chromatin-activating complex (*TAC*) to deposit repressive or activating histone marks, respectively, is well documented. The process of X-Chromosome inactivation (*XCI*), which takes place in placental mammals, results in the suppression of gene expression on an entire female X chromosome due to local epigenetic effects. *XCI* is carried out randomly on one of the two chromosomes in females and is required to balance the expression of genes on the X chromosome between males and females (Figure 9). These epigenetic changes are mediated by the lncRNA *Xist (*X-inactive specific transcript*)*. The ncRNA *RepA* is expressed from within the *Xist* locus and together with *Xist*, they recruit *PRC2*, which catalyses the

deposition of the repressive H3K27me3 chromatin mark across the X chromosome. On the other hand, tight regulation of the XCI process is required to prevent both X chromosomes from being inactivated. *Tsix*, a lncRNA expressed on the sister X chromosome, is able to suppresses *Xist* expression by inhibiting *RepA* recruitment of polycomb in only one of the X chromosomes in females. Both lncRNAs are important for viability and deletion of either *Xist* or *Tsix* lncRNA have shown to be lethal in mice (Marahrens *et al.,* 1997; Sado *et al.,* 2001). The study of dosage compensation has led many groups to explore this phenomenon in greater detail and it has taken a decade of research to decipher the mechanisms. Today, we understand that besides the recruitment of the PRC2 complex, *Xist* itself is highly methylated containing N6-methyladenosine (m6A) residues and the recruitment of RNA-binding motif proteins such as RBM15 is essential for *XIST*-mediated gene silencing (Patil *et al*., 2018). Besides the role in X inactivation, *Xist* have been shown to be associated with a number of cancers with the reactivation of the second X chromosome (Yildirim *et al.,* 2013).

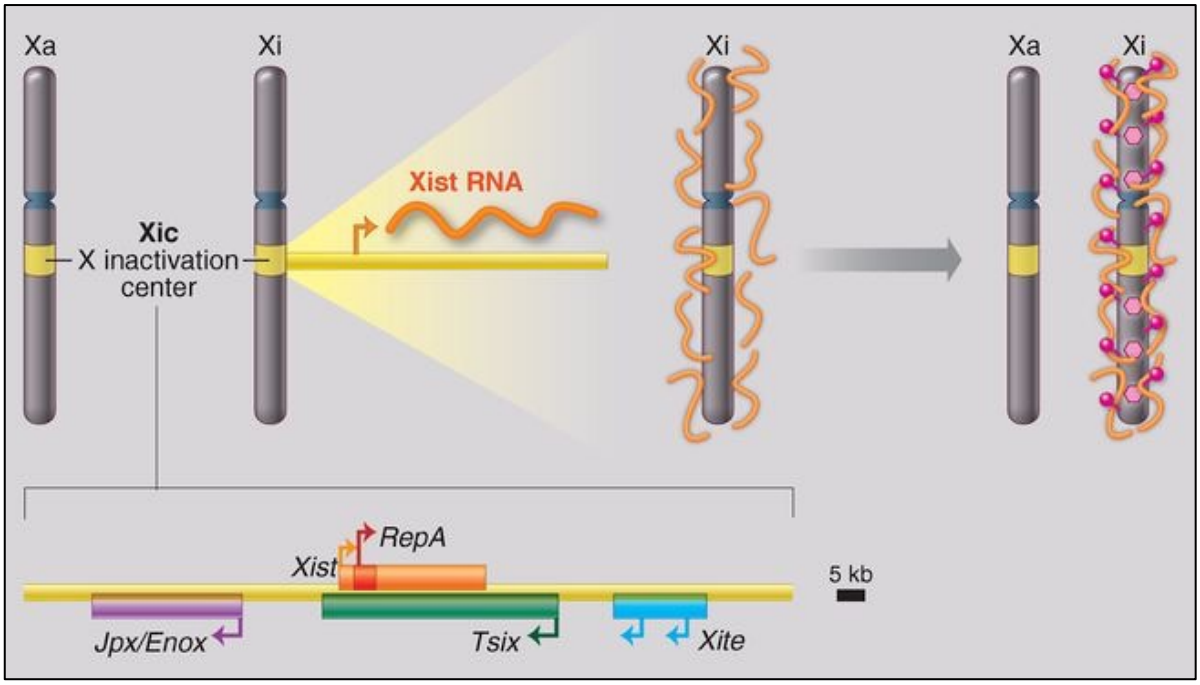


Figure 9: *Xist*, a well-studied lncRNA that is involved in chromatin remodelling.
The process of X chromosome inactivation is orchestrated by Xist. The lncRNA interacts with proteins in the PRC complex to impart repressive histone marks such as H3K27 trimethylation on target genes, which results in the inhibition of gene expression at that locus.
[Image modified from Lee, 2012].

LncRNAs have also been found to induce transcriptional gene silencing in plants through the recruitment of the PRC2 complex. Flowering plants have developed the ability to respond to environmental cues so as to flower at the right time of the year. The promotion of flowering by vernalisation in *Arabidopsis* is regulated by PRC2 and two lncRNAs, *COLDAIR* and *COOLAIR,* on the floral repressor, *FLOWERING LOCUS C* (*FLC*). During winter, the

antisense lncRNA *COOLAIR* is highly expressed and with the processing of the 3′ end of FLC triggers silencing of the sense FLC transcript and promotes the PRC occupancy (Swiezewski *et al*., 2009). An additional layer of gene silencing is carried out by the *COLD-ASSISTED INTRONIC NONCODING RNA (COLDAIR)* that is transcribed from the first intron of the *FLC* locus when *FLC* is repressed. In comparison to *COOLAIR*, lncRNA *COLDAIR* directly interacts with the PRC2 complex and increases the PRC2-mediated H3K27me3 histone mark on the *FLC* gene including the promoter region (Heo and Sung, 2011; Heo *et al.,* 2013). The act of transcription termination by *COOLAIR* and the recruitment of PRC2 to propagate repressive chromatin marks by *COLDAIR* completely silences the *FLC* locus.

LncRNAs are capable of guiding chromatin modifiers in a trans-acting manner to mediate epigenetic modification on specific loci across the genome (Rinn et al., 2007; Tano and Akimitsu, 2012). Since the discovery of *Xist*, a number of lncRNAs that interact with the PRC2 complex to mediate gene silencing have been identified (Khalil *et al*., 2009; Nagano *et al*., 2008; Swiezewski *et al*., 2009; Mohammad *et al*., 2011). Hox transcript antisense intergenic RNA (*HOTAIR*) is one such lncRNA that is expressed from the HOXC cluster and recruits PRC2 to induce repressive histone marks on the HOXD cluster in multiple tissues (Li et al., 2013). The PRC2 complex was identified to associate with *HOTAIR* using an RNA-protein interaction assay and is required for the proper localization the PRC2 across the HOXD cluster. Mutants of *HOTAIR* undergo gene de-repression on the HOXD cluster and display several phenotypes such as homeotic transformation and malformation of the skeletal structure (Li et al., 2013). *HOTAIR* was also shown to be associated with cardiovascular diseases and a number of human cancers (Lai *et al.,* 2017; Gupta *et al.,* 2010; Li *et al.,* 2013).

Although epigenetic silencing is considered one of the major process lncRNAs are involved in, it is not the only form of epigenetic control of gene transcription by lncRNAs. Improved sequencing technologies have shed some light on the presence of a less defined class of lncRNA known as promoter-associated lncRNA (pa-RNAs) that bind to protein factors and target promoters of coding regions (Wang *et al.,* 2008; Song *et al.,* 2012). An example of pa-RNA is lncRNA-CCND1/Cyclin D1, which is expressed upstream of the CCND1 promoter during DNA damage. This lncRNA transcript binds to the Cyclin D1 promoter and modulates the activity of the RNA-binding protein TLS. Alterations in TLS protein prevent histone acetylation of the CREB binding protein as well as p300 and consequently suppresses the expression of Cyclin D1 protein, which is required for cell cycle progression (Wang *et al.,* 2008). LncRNA HOXA transcript at the distal tip (*HOTTIP*) is

another example of a cis-acting lncRNA that mediates gene activation. It is transcribed from the *HOXA* locus and promotes HOXA activation through its interaction with the WRD5/MLL histone methyltransferase (HMT) complex. The loss of *HOTTIP* results in lumbosacral homeotic transformation in mice and a number of studies have also demonstrated that *HOTTIP* plays a major role in genome maintenance and is associated with a number of human cancers (Li *et al.,* 2016; Sun *et al.,* 2017; Zou *et al.,* 2018; Sun *et al.,* 2018).

Transcriptional Regulators

Besides interacting with epigenetic factors and chromatin modifiers, lncRNAs have been shown to regulate gene expression via other processes. LncRNAs can influence transcription directly and prevent DNA binding by transcription factors via several mechanisms such as inhibiting or mediating the recruitment of RNA polymerase II, transcription factors, and/or cofactors to gene promoters. Studies have shown that the growth arrest-specific transcript 5 (*Gas5*) lncRNA acts as a decoy and competes with glucocorticoid response elements for binding to the DNA-binding domain of glucocorticoid receptors (Kino *et al.,* 2010). Gas5 is a starvation-associated repressor and have been shown to be associated with promoting apoptosis (Mourtada-Maarabouni *et al.,* 2008; Kino *et al.,* 2010; Liu *et al.,* 2018). The lncRNA *Lethe* has also been shown to prevent the NF-κB-driven inflammatory responses in mouse fibroblasts by binding with the RelA subunit and preventing it from binding to DNA and thus promoting anti-inflammatory responses (Minton, 2013; Rapicavoli *et al.,* 2013). The lncRNA *Jpx* is an important player in XCI besides lncRNA *Xist*. JPX is a trans-acting lncRNA found to be expressed on the active X chromosome. It is able to upregulate Xist expression on the inactive X allele by binding to CCCTC-binding factor (*CTCF*) and removing it from the inactive X locus (Tian *et al.,* 2010; Sun *et al.,* 2013; Carmona *et al.,* 2018).

Two lncRNAs have been shown to have a role in the nervous system by controlling the fates of neural cells. The interaction between lncRNA rhabdomyosarcoma 2-associated transcript (*RMST*) and the transcription factor *SOX2* is necessary to promote specific neuronal-related gene expression programs (Ng *et al.,* 2013; Wang *et al.,* 2017). Here, *RMST* was found to be a brain-specific lncRNA and modules neural differentiation. *RMST* co-regulates SOX2 target genes upon binding and the association of SOX2 with the regulatory regions of neurogenic genes such as *DLX1, ASCL1, HEY2, and SPS* has been shown to be greatly reduced upon knockdown of *RMST*, and thus inducing changes in the gene profiles of

the neural stem cells (Ng *et al.,* 2013). The second lncRNA is an ultra-conserved ncRNA known as *TUNA/Megamind* that regulates gene expression in a similar fashion as *RMST.* Three RNA-binding proteins (NCL, PTBP1 and hnRNP-K) have been shown to interact with *TUNA* and activates neural genes in differentiating mouse ESCs such as *SOX2* (Ulitsky *et al.,* 2011; Lin *et al.,* 2014).

Functions of post-transcriptional lncRNAs

Examples of lncRNAs that act as post-transcriptional gene regulators have emerged in recent years (Yoon *et al.,* 2014; He *et al.,* 2019). Compared to lncRNAs that execute their function in the nucleus, the roles of many cytoplasmic lncRNAs only began to surface recently. Some of the post-transcriptional roles played by lncRNAs in the cytoplasm are mRNA stabilisation, regulation of target mRNA translation and serving as competing endogenous RNAs (ceRNAs). Many RNA-binding proteins (RBP) and microRNAs were known to regulate the stability of mRNAs and influence the half-life of the mRNA transcript. In recent years, lncRNAs have emerged as another class of molecules that possessed the ability to perform a similar task. The lncRNA known as antisense, longer base pairing of β amyloid-cleaving enzyme (*BACE1-AS*) transcript that is expressed from the BACE1 locus had been found to increase the stability of the BACE1 mRNA for translation by competing with miR-485-5p (Faghihi *et al.,* 2010). The formation of an RNA-RNA duplex between *BACE1* mRNA and lncRNA-*BACE1*-AS mask the binding site that is recognised by miR-485-5p and prevents the miRNA from carrying out gene silencing, which has broad implications in Alzheimer's disease (Faghihi *et al.,* 2008; Hu *et al.,* 2018).

A different class of lncRNAs known as half STAU1 binding site RNA (*1/2-SBS1RNA*) interacts with mRNA transcripts directly and altering them for recognition and degradation by Staufen 1 (*STAU1*) proteins (Kim *et al.,* 2007; Ravel-Chapuis *et al.,* 2012). The mRNA transcripts that would be subject to mRNA degradation via the STAU1-mediated mRNA decay (SMD) pathway would contain a double-stranded RNA motif within the 3′UTR of translationally active mRNA that encodes an ADP-ribosylation factor 1 (*ARF1*) (Gong and Maquat, 2013; Park and Maquat, 2013). A proportion of mRNA transcripts have been found to lack this stem loop structure and would not be recognised by *STAU1* protein. Interestingly, some of the active mRNA transcripts contained an Alu element that allowed imperfect base-pairing with the Alu element of the *1/2-SBS1RNA*, resulting in a decrease in the overall half-life of the mRNA transcript by triggering the formation of a double-stranded RNA binding

motif for the *STAU1* and mRNA decay factor *UPF1* complex (Kim *et al.,* 2005). Conversely, the terminal differentiation-induced ncRNA (*TINCR*) stabilises mRNA through the binding of the 'TINCR box' motif with the target mRNA. Similar to *1/2-SBS1RNA*, *TINCR* was found to be bound to STAU1 but in this case, it aids in the stabilisation of the target mRNA transcript (Kretz *et al.,* 2013). The *STAU1-TINCR* complex was found to stabilise a range of mRNAs that drive somatic tissue differentiation. The association of 1/2-SBS1RNA or lncRNA *TINCR* with mRNA transcripts do not trigger SMD but changes the overall stability profile of the mRNA. Recent studies have shown that the STAU proteins are considered to be multifunctional and able to stabilise and destabilise target mRNA (Heraud-Farlow and Kiebler, 2014; Heraud-Farlow *et al.,* 2013). The actual mechanism remains obscure and further investigation is needed to unravel the other conditions that influence the mRNA turnover upon lncRNA binding.

Besides influencing the stability of mRNAs, lncRNAs were found to be repressors (e.g. *lincRNA-p21, BC1*) or activators (e.g *AS Uchl1*) of target mRNA translation. The dynamics between human antigen R (*HuR*) and lincRNA-p21 in the cell influenced mRNA translation of target genes (Tang *et al.,* 2015; Yang *et al.,* 2016). LincRNA-p21, also known as Tumour protein p53 pathway corepressor 1 (Trp53cor1), was found to be an important translation modulator highly enriched in the cytoplasm. In an experiment using RNA immunoprecipitation (RIP) to search for RNA binding with the RNA-binding protein *HuR,* lincRNA-p21 was found to be one of the interactors in the pool. In the presence of *HuR*, lincRNA-p21 have been shown to be degraded rapidly (Yoon *et al.,* 2012; Wilusz and Wilusz, 2012). *HuR* acts as an mRNA stabiliser and promotes CTNNB1 and JUNB mRNAs translation. In the absence of *HuR*, lincRNA-p21 transcript levels are elevated and bound to CTNNB1 and JUNB mRNAs (Turner *et al.,* 2014). This base-pairing event further recruits the translation repressor Rck to prevent translation of target mRNAs and also increased the ribosome drop-off rate (Yoon *et al.,* 2012). The rodent brain cytoplasmic 1 (*BC1*) lncRNA is a brain-specific transcript acting as a translational repressor through its interaction with eukaryotic translation initiation factor eIF4A and poly(A)-binding protein (PABP) (Muddashetty *et al.,* 2002; Wang *et al.,* 2005). The transport of *BC1* to the dendrites of neurons would thus influence the local translation processes and the loss of *BC1* was shown to result in neuronal excitability and susceptibility to epileptic seizures (Zhong *et al.,* 2009; Gitaí *et al.,* 2009). On the other hand, the expression of dopaminergic neuron-specific lncRNA known as antisense Ubiquitin carboxy-terminal hydrolase L1 (*AS Uchl1*) promotes cap-independent translation of *Uchl1.* The repetitive domain SINEB2 that was found to be

embedded in the *AS Uchl1* recruits polysomes and was found to be essential for promoting translation (Carrieri *et al.,* 2012; Podbevšek *et al.,* 2018).

The search for lncRNAs that interact with other ncRNAs transcripts such as miRNAs led to the discovery of a group of ncRNA known as ceRNAs. As the name suggested, these group of lncRNA transcripts lncRNA stabilise mRNA levels by serving as a molecular 'decoy" for miRNAs or proteins that targets them (Poliseno *et al.,* 2010; Salmena *et al.,* 2011). The 3′-UTR region of muscle-specific, long intergenic non-coding RNA, muscle differentiation 1 (*LINCMD1*) harbours miRNA-binding sites for miR-133 and miR-135 and thereby sequester these miRNAs away from their mRNA targets. The reduction in miRNA availability positively upregulates the expression of transcription factors such as Mastermind-like 1 (*MAML1*) and Myocyte enhancer factor 2C (*MEF2C*) and thereby drives muscle-specific gene expression programs (Cesana *et al.,* 2011). Increasing evidence has pointed towards lncRNA-miRNA interactions as contributors to several cancers such as breast, prostate, gastric, lung and liver cancer. Some of the well-known lncRNAs were also found to attenuate miRNA regulation. The binding of *HOTAIR* to miR-331-3p relives the repression of human epidermal growth factor receptor 2 (*HER2*) by the miRNA and drives the progression of gastric cancer (Liu *et al.,* 2014). Similarly, *H19* lncRNA functions as a molecular sponge for miRNAs let-7 and promote cancer cell proliferation and metastasis with the elevation of c-Myc expression (Zhou *et al.,* 2017; Kallen *et al.,* 2013). The lncRNA Human universal load carrier (*HULC*) indirectly stimulate gene expression programs by sequestering miRNA-372 and thus increased the phosphorylation activity of CREB proteins by PRKACB (Protein Kinase CAMP-Activated Catalytic Subunit Beta). The upregulation of *HULC* has been found to be associated with different cancers such as hepatocellular carcinoma and gastric cancer (Wang *et al.,* 2010; Panzitt *et al.,* 2007; Yang *et al.,* 2015; Chen *et al.,* 2017).

1.5 Emerging roles of lncRNAs in *Drosophila*

Genome-wide transcriptional studies indicated that the genome of *Drosophila melanogaster* codes for more than 1,000 lncRNAs, yet the functions of these non-coding RNA molecules remains largely unknown (Brown *et al.,* 2014). The first lncRNA discovered in *Drosophila* was the transcriptional elongation of *bithoraxoid (bxd)* ncRNA (Lipshitz *et al.,* 1987). It is expressed within the HOX gene cluster and positioned adjacent to the *Ultrabithorax (Ubx)* gene. The *bxd* lncRNA functions as a repressor of *Ubx* gene, with mutations resulting in homeotic transformations and the generation of a four-winged fruit fly

(Petruk *et al.,* 2006, Pease et al., 2013). However, subsequent studies have revealed confounding results using different methods of generating *bxd* mutants. It was suggested that it was the act of transcription of this lncRNA that affected the expression the *Ubx* gene rather than the lncRNA itself. The inversion of the promoter region had little effect on the expression of *Ubx* even though transcription was occurring in the other direction (Pease *et al.,* 2013). Similarly, a Cdx-like gain of function of *Ubx* was observed when the promoter region of *Ubx* was deleted (Sipos *et al.,* 2007). This has prompted the need for more in depth studies and careful analysis of lncRNA functions and the need to be cautious when drawing conclusions from loss of function studies.

Other well-studied *Drosophila* lncRNAs are *roX1*, *roX2* and heat shock RNA omega (*hsr ω*). Both *roX1*and *roX2* lncRNAs were found to be involved in the dosage compensation process of the X chromosome in male flies and implicated in epigenetic gene regulation through histone modification (Khalil *et al.,* 2009; Ilik and Akhtar, 2009). These two lncRNAs interact with the Male-specific lethal (*MSL*) complex to acetylate histones on the X chromosome, leading to hyperactivation of transcription. *Hsr ω* is transcribed from the heat shock response hsr-omega locus in response to cellular stress and has been detected at high levels in many tissues, particularly in the nucleus (Bendena *et al.,* 1991). Three long transcripts are generated during transcription from the hsr-omega locus by RNA pol II after undergoing alternative splicing and polyadenylation. These lncRNA transcripts were found to be regulating apoptosis and have been implicated in coordinating stress responses and recruiting various RNA-binding proteins into omega speckles upon stress induction (Mallik and Lakhotia, 2009; Singh and Lakhotia, 2015). The majority of fly strains carrying *hsr ω* mutations are embryonic lethal (Jolly and Lakhotia, 2006).

A number of studies have found that lncRNAs also play a role in *Drosophila* behaviour and cognition. For example, the lncRNA *bereft* (*bft*) is expressed in the extrasensory organ and was shown to be vital for bristle formation (Hardiman *et al.,* 2002). Another lncRNA known as *CASK regulatory gene* (*CRG*) is a cis-acting ncRNA transcript that positively regulates its neighbouring gene *CASK* (*Ca2+/calmodulin-dependent protein kinase*). Loss of lncRNA *CRG* has been found associated with defects in *Drosophila* locomotor activity and climbing ability (Li *et al.,* 2012). Two lncRNAs have been found to influence the behaviour of the fruit fly. *Sphinx* lncRNA was shown to associate with male courtship behaviour whereas lncRNA yellow-achaete intergenic RNA (*yar*) was found to have an impact on the circadian rhythm and is important for proper sleep regulation (Wang *et*

al., 2002; Soshnew *et al.,* 2011). Apart from these few examples, we know relatively little about lncRNAs in *Drosophila*.

1.6 Tools to understand lncRNAs functions

The application of single-molecule RNA fluorescence in situ hybridization (smFISH) has been a molecular biology technique used in the field for many years (Gall JG and Pardue, 1969). This method was commonly employed to reveal the subcellular localisation and expression pattern of lncRNAs so as to have an insight towards the role the play in the cell (Levsky JM and Singer, 2003; Dunagin *et al.,* 2015; Orjalo and Johansson, 2016). The innovations in smFISH technologies in recent years have increased the sensitivity of the assays and allowed the detection and quantification of lncRNA transcripts that have been known to be low-abundance (Wang *et al.,* 2012; Jandura *et al.,* 2017; Choi *et al.,* 2018). Using RNA fluorescence in situ hybridization methods, two cis-acting lncRNAs called RNA on the X 1 and 2 (roX1 and roX2), were found to be expressed from the male X chromosome and were shown to be involved in the dosage compensation process (Meller and Rattner, 2002). The use of RNA fluorescence in situ hybridization coupled with immunofluorescence subsequently revealed that the roX RNAs co-localise with the subunits found in the male-specific lethal protein complex that is made up of five proteins. The assembly of this complex was essential for the spreading of the roX RNAs and hyperactivation of the X chromosome (Franke and Baker, 1999; Kelly *et al.,* 1999; Meller and Rattner, 2002; Oh *et al.,* 2003).

Interestingly, due to an overlap in function, the male sterility phenotype requires the deletion of both lncRNAs and either lncRNA was found to be functionally redundant (Stuckenholz *et al.,* 2003). Genome engineering methods such as recombination-TALEN/CRISPR technologies to target lncRNA genes have allowed the perturbation of lncRNAs of interest to investigate their function. Various methodical advances have greatly improved the in vivo functional characterization of lncRNAs. The study of lncRNA function in vivo is necessary as it has been shown in some lncRNA mutants that the mutation had little or no phenotype during development compared to the in vitro studies performed on the same lncRNA (Oliver *et al.,* 2015; Sauvageau *et al.,* 2013; Mattick, 2003; Grote *et al.,* 2013). An example of such a lncRNA is *MALAT1* that have been found to be highly abundant in vertebrate cells. *MALAT1*was found to be localised to the nuclear speckles and have been shown to have important roles in cancer metastasis and synapse formation in vitro (Gutschner *et al.,* 2013; Schmitt and Chang, 2016). However, *MALAT1* KO mouse models were found to

be viable and fertile with no apparent developmental abnormalities (Eißmann *et al.,* 2012; Nakagawa *et al.,* 2012; Zhang *et al.,* 2012).

In recent years, various RNA-centric and protein-centric methods have been developed to probe for lncRNA interactions with DNA, RNA and protein molecules (König *et al.,* 2012, West *et al.,* 2014; Chu *et al.,* 2015; Yang *et al.,* 2015; Cheetham and Brand, 2018). Mechanistic studies to identify lncRNA binding partners have helped researchers discern the pathways and understand the mode of action of the lncRNA of interest. Through their interaction networks, we can gain a better insight into the diseases that may be associated with lncRNA dysregulation. Chromatin isolation by RNA purification (ChIRP), chromatin isolation by RNA purification (CHART) and RNA antisense purification (RAP) are some of the methods widely used in the field to investigate lncRNA binding sites in the genome (Chu *et al.,* 2011; Quinn *et al.,* 2014; Engreitz *et al.,* 2015). Using *Drosophila roX* RNAs as an example, it was shown that dosage compensation is carried out by roX RNAs and the male-specific lethal protein complex (Georgiev *et al.,* 2011). The MSL complex encompasses the maleless (MLE) subunit, male-specific lethal 1 to 3 (MSL1 to MSL3) subunits and males absent on the first (MOF) subunit (Bhadra *et al.,* 1999; Bhadra *et al.,* 2005). For a long time, the binding sites on the X chromosome by roX RNAs were unknown. Using dChIRP, the authors generated biotinylated oligo probes antisense to the roX RNAs and pulled down regions that were bound to the probes (Quinn and Chang, 2015). This method can be paired with sequencing to look for the regions that were bound by the lncRNA. Similarly, the interacting sites of the lncRNA with the protein subunits was revealed using dChIRP coupled with mass spectrometry. The domains D1, D2 and D3 of the MSL proteins were discovered to be interacting with the roX lncRNAs (Chu *et al.,* 2012). In a different experiment, the CHART method was used to study roX2 lncRNA and found similar binding sites for the lncRNA on the X chromosome with the MSL proteins (Simon *et al.,* 2011).

1.7 *Drosophila* – a model organism for the study of lncRNAs

Drosophila melanogaster, also known as the fruit fly, has been used as a model organism for many biological enquiries as it shares many molecular and cellular features with mammals. Ever since it was introduced as a scientific research tool by Thomas Hunt Morgan in the 1900s, the fruit fly has been used extensively for many ground-breaking genetic studies from mechanistic studies to genetic variation as well as pathway studies (St Johnston, 2002; Bier, 2005). Biologists have used the fruit fly in many additional areas of research such as

developmental biology, cancer, metabolism and behavioural studies (Halder and Mills, 2011; Nichols *et al.,* 2012). It can be considered to be a relatively complex organism but at the same time have an unrivalled ease of usage in the lab. The are many compelling reasons why *Drosophila* is such an amazing and fascinating tool to work with. *Drosophila* can be considered one of the most understood organisms in the lab given its long history. It has a fast generation time of 10.5-11 days at 25 ºC and it is easy to breed in the lab and given its small size it occupies little space. The fruit fly has a compact genome (~180 Mb in size) and it is known that approximately 60-75% of human disease-linked genes have a functional orthologue in *Drosophila*, allowing researchers to use the fruit fly to understand biological functions that are conserved among animals (Reiter *et al.,* 2001). The fruit fly genome has been fully sequenced and a large amount of transcriptomic data and metabolomic data sets were made available from the modENCODE project as reference datasets (Adams *et al.,* 2000; Chintapalli *et al.,* 2007; Chintapalli *et al.,* 2013; Brown *et al.,* 2014). Other positive factors include low maintenance cost, high fecundity and the wide range of genetic tools, gene-specific knockdown and mutant lines that have been generated over the years.

The success of *Drosophila* as a model for genetic studies was made possible with the wide variety of genetic tools that have been created over the century. The development of balancer chromosomes that prevent recombination allow the maintenance of deleterious mutations (Wallace, 1956). The balancer chromosomes are an essential component of mating schemes in genetic crosses even today. They contain multiple inversions that were lethal in homozygotes. The presence of dominant marker mutations ensured that a recessive mutant allele would be kept over the dominantly marked chromosomes. Different balancer chromosomes have been generated and some of the commonly used markers on different chromosomes are *FM7* (X chromosome), *CyO/SM6a* (2nd chromosome) and *TM3*/*TM6B* (3rd chromosome) (Roote and Prokop 2013). The wide range of genome editing methods from chemical mutagenesis to CRISPR/Cas system allow the disruption of genes for functional analysis making *Drosophila* an attractive model to work with (Nasrat *et al.,* 1954, Lin *et al.,* 2015). The availability of RNAi lines for almost every gene have made *Drosophila* a model of choice for many fields of research (Ni *et al.,* 2009; Perkins *et al.,* 2015). A few Nobel prizes have been awarded throughout the years for work performed with the fruit fly. Besides the Nobel prize in Physiology or Medicine that was awarded to Thomas Morgan Hunt in 1933 for “for his discoveries concerning the role played by the chromosome in heredity”, the other noteworthy discovery was the identification of key developmental genes that were involved in the patterning of the body and were also conserved in other organisms (Nusslein-Volhard and

Wieschaus, 1980; Lewis, 1978). For this ground-breaking work, Edward B. Lewis, Christiane Nüsslein-Volhard and Eric F. Wieschaus received the Nobel Prize in Physiology or Medicine for "their discoveries concerning the genetic control of early embryonic development" in 1995. Today, *Drosophila* is commonly used as a genetic model to study many pathological conditions and to understand the physiological mechanisms involved in processes such as ageing, cancer, metabolic diseases and, neurodegenerative disorders (Zou *et al.,* 2000; Baker and Thummel, 2007; Rudrapatna *et al.,* 2012; Lu and Vogel, 2009; Polesello *et al.,* 2011).

According to the Ensembl database, lncRNAs comprise 2,366 of the 17,559 genes in the *Drosophila* genome and make up 13.5 % of the total number of expressed transcripts. Although a wealth of information such as the loci, sequence information, conservation and expression data sets for lncRNAs in *Drosophila* are available, only a small fraction of lncRNAs have been investigated experimentally. The availability of genetic tools such as GAL4-UAS system and CRISPR/Cas9 to functionally characterise lncRNAs makes the fruit fly an excellent model organism to study their function in vivo.

1.8 *Drosophila* life cycle

One of the attractive features of using the fruit fly is its well-studied development from embryo to adult (Figure 10). The fruit fly has a relatively short generation time and the life cycle of a fly is about 10.5-11 days at 25 °C. After the fertilized eggs are laid, embryogenesis takes about 22-24 hrs to complete before hatching into first instar larvae. The embryonic development of *Drosophila melanogaster* consists of 17 stages defined by Volker Hartenstein and José Campos-Ortega and during this period, the blueprint of organs and the appendages of the adult fly are formed (Campos-Ortega and Hartenstein, 1985). *Drosophila* go through three larval stages over a time of 4 days before transforming into pupae. The metamorphosis process takes another 5-6 days before the emergence of an adult fly (Wixon and O'Kane, 2000). During pupation, the larval tissues are degraded to make room for the development of adult organs by the different imaginal disks. There are 10 pairs of imaginal disks that establish the different adult organs and extensive research using these imaginal discs led to an understanding of the reshaping process that takes place during *Drosophila* metamorphosis (Bate and Arias, 1991; Beira and Paro, 2016, Dye *et al.,* 2017). The virgin fruit flies that are newly hatched have physically distinctive features compared to the more mature adults and thus making it easy to isolate and obtain virgin males and females for genetic crosses.

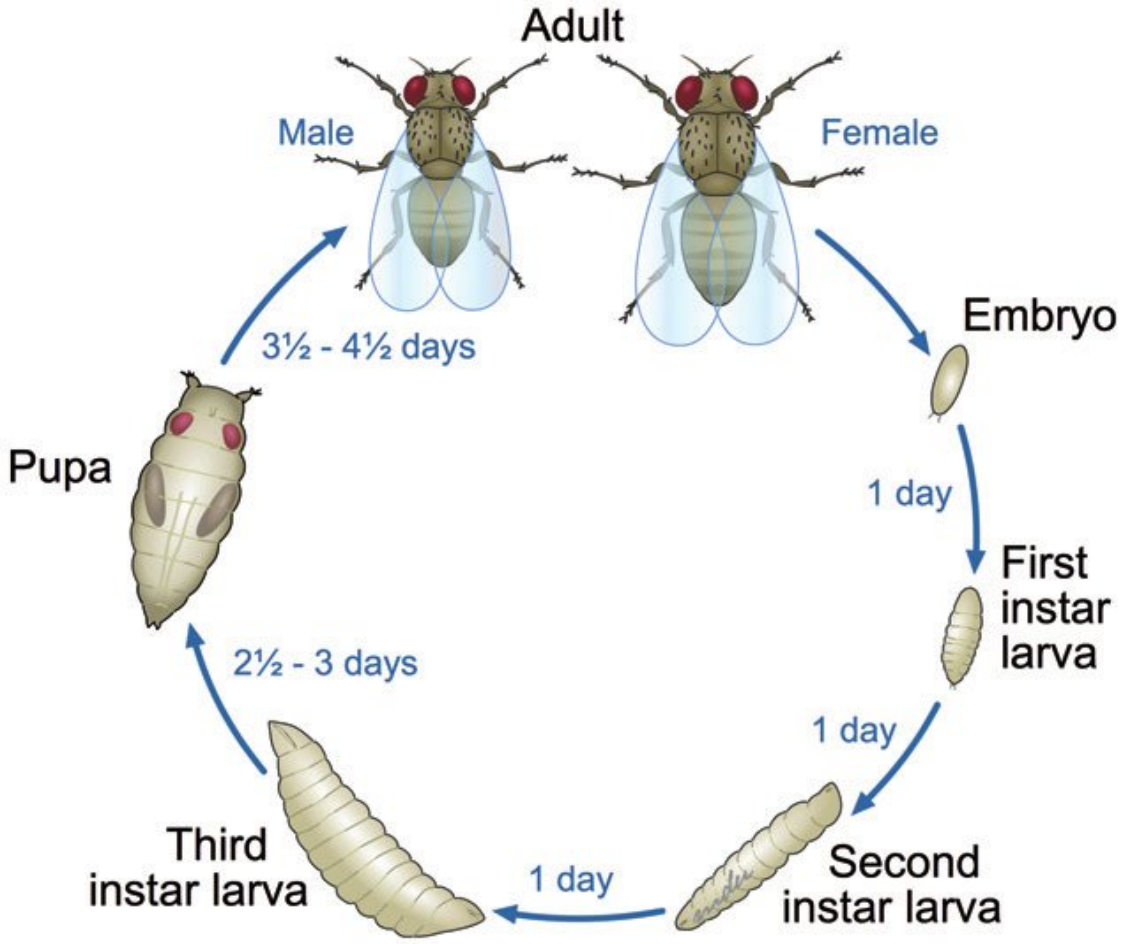


Figure 10. The life cycle of *Drosophila melanogaster*.
After eggs are laid, it takes approximately 10.5-11 days to complete the life cycle at 25 ºC. The *Drosophila* development comprises embryogenesis, larval, pupal and adult stages. During the larval stages, the larvae molt from first instar, to second instar and eventually third instar stage before undergoing pupation. [Image taken from Ong *et al.,* 2014].

1.9 Genetic Manipulations using *Drosophila*

The advances in genome editing techniques have allowed researchers to perform mutagenesis with higher precision and accuracy. The use of conventional methods such as X-rays and chemical mutagens such as ethyl methanesulfonate (EMS) were some of the earlier approaches used to generate mutant flies for loss of function studies (Alderson, 1965; Epler, 1966). However, the mutants generated were random. Hypomorphic alleles with subtle phenotypes make up a substantial portion of the mutants due the EMS-based mutagenesis method that generated single point mutations. It was difficult to trace the mutations as there were no traceable sequences or markers tagged to the lesion. Furthermore, it had to be tested if mutagenesis had occurred on more than one site in the genome (Blumenstiel *et al.,* 2009). The verification process was laborious as complementation assays or whole-genome sequencing were required. P-element-mediated transformation became one of the

revolutionary tools that is still used to generate mutant flies today. The method enables disruption of genes and also the introduction of transgenes into the fly genome (Venken and Bellen, 2007). The P-element is a *Drosophila* transposon that contains an inverted repeat. P-elements have the ability to excise themselves from the genome and re-insert at a different location with the aid of a transposase encoded in the P-element sequence (Huang *et al.,* 2009; Majumdar and Rio, 2015). Since the enzyme is encoded within the P-element sequence, the insertion would not be controllable and would allow the transposon to "jump" around. Therefore, one way of overcoming this problem was to separate the two components of the P element into separate flies and providing the transposase when needed (Rubin and Spradling, 1982). *P-elements* had the tendency to be inserted near actively transcribed genes and this insertion event often disrupts the gene directly or via imprecise excisions after a second translocation event. However, there were caveats when choosing this method. Similar to earlier methods, the nature of *P-element*-mediated transformation in the genome are a random process and require a large amount of time to identify and map the site of insertion in the genome. On top of that, multiple insertion lines would be required for analysis so as to certify that the phenotype observed was not due to an insertion artefact but rather the expression of the transgene itself. Compared to the mutagenesis with either X-ray or EMS, the *P-element* system had a variable transformation effectiveness and a strong tendency to be inserted at "hotspots" whereas the PhiC31 system allows targeted transgenesis which is more efficient. (Bateman *et al.,* 2006).

Zinc-finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) are two important tools that alter a genomic site by inducing DNA double-stranded breaks (DSBs) and activating the DNA repair mechanism of the cell to repair the genomic loci via non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Gaj *et al.,* 2013). Both are artificial chimeric enzymes containing a *Fok1* DNA-cleavage domain as well as a DNA binding domain that has been engineered to recognise a specific sequence. ZFNs and TALENs can also be merged with other enzymes such as nucleases and DNA-histone methyltransferases which allow researchers to study effects on the genomic structure and function. The disadvantage of ZFNs and TALENs is the DNA-binding affinity and specificity of the designed proteins as these factors are important determinants of the capacity of the enzymes to modify the gene of interest. These tools are rarely used as it is difficult to clone and they are often not successful in modifying the genome.

The GAL4/UAS expression system (Brand and Perrimon, 1993) is an important genetic tool used in *Drosophila melanogaster* to express a gene of interest in a time-dependent manner with tissue specificity. The system is made up of two components to drive the expression of a particular gene. The yeast transcriptional activator domain Gal4 is placed under the control of a *Drosophila* enhancer/promotor in one fly whereas the 'Upstream Activating Sequence' (UAS) followed by the gene of interest is placed in a separate fly line. When the two flies are mated, Gal4 binds to the UAS sequence, leading to transcription of the gene of interest. In this binary system the promoter driving Gal4 determines the expression pattern in which the gene of interest downstream of the UAS sites is going to be expressed. Gal4 drivers that are ubiquitous (*Actin*5c), tissue-specific (pan- neuronal *elav*), or those inducible by heat-shock have allowed us to perform targeted gene expression.

Clustered regularly interspaced short palindromic repeats (CRISPR) is currently one of the most popular genome-editing tools that is used in a wide range of model organisms (Cho *et al.,* 2013; DiCarlo *et al.,* 2013; Gratz *et al.,* 2013; Hwang *et al.,* 2013). Compared to the use of ZFN/TALEN that would require the engineering of proteins for the specific gene target; the use of a variety of single guide RNA (sgRNA) allows the Cas9 protein to target different regions of the genome. Briefly, CRISPR/Cas9 allows the user to easily target and perform site-specific gene modifications using an sgRNA to target the DNA of interest (Port *et al.,* 2014; Gaj *et al.,* 2013). The only requirement is that the sgRNA sequence in the genome is directly adjacent to a NGG sequence known as the protospacer adjacent motif (PAM) that is required for recognition by Cas9 proteins (Jinek *et al.,* 2012; Cong *et al.,* 2013). Selection of the sgRNA plays an important part in the editing efficiency of Cas9 and since the introduction of the CRISPR/Cas9 system, several studies have been performed to improve the efficiency of selected sgRNAs (Liang *et al.,* 2017; Meier *et al.,* 2017, Erard *et al.,* 2017). The sgRNA guides the Cas9 protein to the target site to induce a double stranded DNA break that is then repaired by the DNA repair machinery using either homology-directed repair (HDR) or non-homology directed repair (NHEJ) pathways (Glaser *et al.,* 2016). The "knock-in" process of co-injecting a donor vector carrying different selection markers such as 3xP3-RFP with the guide RNA plasmid allow HDR pathway to take place and accelerates the screening process. On the other hand, the study of structure-function can also be performed using the NHEJ repair pathway that generates point mutations in the endogenous locus.

1.10 Characterising lncRNA functions in vivo: the pros, the cons and the challenges

With the expansion in the number of lncRNAs discovered, it is tempting to believe that many of these play some important biological roles (Consortium *et al.,* 2007; Lander *et al.,* 2001). Indeed, lncRNAs have been suggested to carry out a plethora of biological functions in the cell and careful experimental dissection for each of this ncRNA is needed to differentiate them from transcriptional noise (Mattick, 2009). One of the concerns pertaining to the study of gene function in vivo is that it is much more time-consuming and therefore expensive to generate lncRNA mutants as compared to knocking them down or out in cultured cells. Although the function of several lncRNAs has been revealed by in vitro models (Bernard *et al.,* 2010; Cabezas-Wallscheid *et al.,* 2014; Zhang *et al.,* 2009), there is limited genetic evidence to support their in vivo functions.

Animal models with simple biological complexity such as *Caenorhabditis elegans* (nematode)*, Drosophila melanogaster* (fruit fly)*, Danio rerio* (zebrafish) have been widely used for genetic screens to understand biological processes and they are well suited for the physiological study of lncRNAs (Kapusta and Feschotte, 2014; Ulitsky, 2016). Gain-of-function and loss-of-function genetic models are common strategies used to investigate the roles of protein-coding genes and have been employed to understand some of the lncRNAs discovered earlier such as *Xist* and *AIRN* (Marahrens *et al.,* 1997; Sleutels *et al.,* 2002). A small fraction of lncRNAs have been tested in a similar manner subsequently and many of them have been discovered to be dispensable (Zhang *et al.,* 2012; Sommerkamp *et al.,* 2019; Goudarzi *et al.,* 2019; Schor *et al.,* 2019). Despite the in vivo results observed, outcomes obtained can sometimes be inconsistent with cell-based studies and extra care must be taken to interpreting both positive and negative results (Bassett *et al.,* 2014; Li and Chang, 2014; Boettcher and McManus, 2015).

Functional studies have been widely conducted by attenuating or ablating the expression of a gene of interest. Each of these methods has its own advantage and drawback and thus a combination of approaches is required to validate the observed effect (Figure 11). Genetic tools can be used to target lncRNA transcripts at the RNA level such as the introduction of morpholinos or RNA interference (RNAi) or at the DNA level through the manipulation of lncRNA genomic loci (Wilk *et al.,* 2016; Chen, 2016). To functionally characterize a lncRNA, an important point to note is that the combination of several methods

rather than a single approach is required to fully characterise a lncRNA locus (Goff and Rinn, 2015).

shRNA, siRNA or dsRNA knockdown

RNA-targeting techniques are frequently used to knockdown gene expression through the introduction of siRNA, shRNA or morpholinos (Guttman *et al.,* 2011; Flynn *et al.,* 2015; Goudarzi *et al.,* 2019). RNA-targeting methods degrade the lncRNA transcript itself and allow the study of the functions mediated by the RNA. Tissue-specific suppression of lncRNA *Gomafu/MIAT/Rncr2* in mouse models using shRNA has revealed a role of this lncRNA in the mediation of anxiety-like behaviour and the development of neuropsychiatric disorders (Rapicavoli *et al.,* 2010; Spadaro *et al.,* 2015;). Likewise, the use of shRNA-mediated knockdown of lncRNA AT-rich interactive domain 2-IR (*Arid2-IR)* in mouse has unveiled a potential therapy target for renal inflammatory disease (Zhou *et al.,* 2015; Yang *et al.,* 2019). Besides mouse models, zebrafish is a well-known model for vertebrate studies and thousands of lncRNAs have been found to be expressed at various stages during its embryonic development. The injection of morpholinos antisense oligos (MO) into zebrafish is an alternative knockdown approach that is independent of the RNAi machinery (Ulitsky *et al.,* 2011; Goudarzi *et al.,* 2019). These synthetic molecules target splice junctions or conserved regions and can be administer into zebrafish embryos when they are still at the one cell stage to knockdown lncRNA expression throughout development. *Megamind* and *Cyrano* are 2 examples of lncRNAs whose function has been elucidated using MO. Attenuation of these lncRNAs has been shown to disrupt the nervous system development and affect proper embryonic development (Ulitsky *et al.,* 2011; Lin *et al.,* 2014).

These use of RNA-targeting methods is not without limitations. The advantages of using RNAi is allowing the user to conditionally control gene expression in a temporal or spatial manner and also the ability to perform a high throughput screening of a list of lncRNAs to determine their physiological function. However, the relative ability of RNAi or antisense oligos to suppress a lncRNA target is dependent on its subcellular location. Many lncRNAs are found to be confined to the nucleus and some of them have been found to regulate gene expression in a cis-acting manner (Derrien *et al.,* 2012; Lennox and Behlke, 2016; Kopp and Mendell, 2018). Besides the potential off-target and toxicity effects in using both siRNA or shRNA, it remained a challenge to effectively knockdown lncRNAs that are chromatin associated or localised to the nucleus (Zeng and Cullen, 2002; Vickers *et al.,*

2003). As the RNAi machinery is located mainly in the cytoplasm, the general efficiency in depleting nuclear-retained lncRNAs is limited (Zong *et al.,* 2015).

Alternative methods that uses steric interfering RNA analogs such as locked nucleic acids (LNAs) and morpholino have been shown to target nuclear lncRNAs but these synthetic oligos cannot be encoded genetically and instead have to be injected or supplied externally (Ulitsky *et al.,* 2011; Qu *et al.,* 2016). Due to the mode of delivery, performing high-throughput screening is possible with RNAi but not with these analogs. Another consideration for the use of both methods is its knockdown efficiency as sufficient levels of knockdown is required to repress the lncRNA transcript for a phenotypic effect to be observed (Atkinso *et al.,* 2018). Furthermore, some studies have compared the use of gene editing methods and morpholino knockdowns in zebrafish. From the gene editing results, it is observed that a higher rate of non-specific phenotypes in found in morpholino knockdowns compared to CRISPR/Cas9 gene editing technique (Novodvorsky *et al.,* 2015; Place *et al.,* 2017; Kok *et al.,* 2014).

Genetic deletion

DNA-targeting approaches that manipulate the genomic loci have been widely used to study lncRNA mutants in vivo. One of the shortcomings for the use of RNA-based method is its inability to distinguish cis-acting from trans-acting lncRNAs and this issue can be circumvented with the deletion of the lncRNA itself (Figure 11). Gene deletion methods are strong proof of functionality, a lncRNA locus can be considered to be dispensable with the absence of an observable phenotype. However, caution should be placed when interpreting these results as an absence of overt phenotype does not necessarily dictate an absence of function. Transgenic mutants defective for a lncRNA of interest can be generated with the disruption at the germline level compared to RNA-targeting methods. In addition, the deletion of the whole lncRNA locus can be combined with other methods such as the introduction of a reporter gene for phenotypic investigation (Valenzuela *et al.,* 2003; Chen *et al.,* 2017). Replacement of the lncRNA with a fluorescence protein construct enables the expression profiling of the lncRNA of interest and have been used to study 18 mice lncRNAs (Sauvageau *et al.,* 2013). LncRNA *Peril, Mdgt, Pantr2* are some examples of lncRNAs which had their tissue expression patterns elucidated using KO models (Sauvageau *et al.,* 2013).

LncRNA *MALAT1* and lncRNA *C130071C03Rik/LINC00461* are examples of lncRNA knockouts that have been found to be more context-specific and important under certain physiological conditions and not required for normal development (Zhang *et al.,* 2012; Nakagawa *et al.,* 2012; Eißmann *et al.,* 2012; Oliver *et al.,* 2015). *MALAT1* is a well-known lncRNA that is conserved, highly abundant and localises to the nuclear speckles (Tripathi *et al.,* 2010). It was hypothesized that *MALAT1* could possess critical biological functions for normal physiological development as it was observed that the lncRNA plays an important role in cell proliferation, cancer invasion and metastasis (Zhou *et al.,* 2015; Yang *et al.,* 2015; Arun *et al.,* 2016). However, in three separate loss-of-function genetic studies, this ncRNA was surprisingly found to be dispensable for the viability of the animal. Furthermore, the dysregulation of *MALAT1* in these knockout mice did not display any changes in splicing factor levels (Zhang *et al.,* 2012; Nakagawa *et al.,* 2012; Eißmann *et al.,* 2012). Likewise, previous studies have indicated that lncRNA *C130071C03Rik/LINC00461* contributes to cell proliferation and migration in glioma cells (Oliver *et al.,* 2015). However, a lack of overt phenotype has been observed with the deletion of the genomic locus of this lncRNA (Oliver *et al.,* 2015). Even though no obvious defects were observed in *MALAT1* knockouts in normal physiological conditions, crossing of these KO mutants with the MMTV-PyMT human breast cancer mouse model led to a decrease in metastatic burden and suggested a role for lncRNAs during cellular responses to stress or disease conditions such as cancer progression (Arun *et al.,* 2016; Kim *et al.,* 2018; Wang *et al.,* 2019).

Integration of a premature polyadenylation cassette

During the removal of a whole DNA locus, underlying regulatory elements close to the lncRNA genomic locus may contribute to the observed phenotype when they are deleted. Therefore, the use of multiple independent genetic models allows proper separation of the effects due to the loss of the lncRNA from effects due to the loss of underlying DNA elements (Bassett *et al.,* 2014; Kopp and Mendell, 2018). The insertion of transcriptional terminators is an elegant approach that serves as a complementary method to whole gene ablation mutagenesis (Grote *et al.,* 2013; Engreitz *et al.,* 2016). LncRNA mutants with premature transcriptional termination can be generated by introducing multiple poly adenylation (poly-A) signals close to the 5’ end of the lncRNA locus (Sleutels *et al.,* 2002; Bond *et al.,* 2009). Using the *Fendrr* lncRNA as an example, both methods have been applied to the genomic locus of this ncRNA and a reduction in viability was observed in both instances (Grote et al., 2013; Sauvageau *et al.,* 2013). However, when using the polyA

termination method, an embryonic lethal phenotype was seen in the in vivo study performed by Grote *et al.,* 2013 whereas a perinatal lethality was observed in mouse mutants generated by Sauvageau *et al.,* 2013 that featured the deletion of the lncRNA genomic locus. Even though a similar phenotype was observed, the authors were able to distinguish between the DNA element from the RNA product itself. The lncRNA was found to be the functional element in this case as the phenotype for the *Fendrr* mutant was rescued using a transgene without its functional *Foxf1* neighbour (Grote *et al.,* 2013). The use of rescue experiments allows proper differentiation between effects caused by the loss of DNA elements and loss of the lncRNA itself. DNA fragments containing the lncRNA can be introduced back into the system by using bacterial artificial chromosomes (BACs) or cDNA transgenes or by injection of the RNA transcript back into the organism (Flynn and Chang, 2015; Goff and Rinn, 2015).

Inversion

The use of inversion is a method whereby the direction of transcription is changed and while retaining the DNA elements at the lncRNA genomic loci. It is a less commonly used methodology compared to the other two approaches mentioned above. Nonetheless, valuable information can be obtained with inversion of the promoter or whole lncRNA locus as exemplified by studies of the *Drosophila bxd* lncRNA. As with RNA-targeting methods, there are caveats also with the use of DNA-targeting techniques. While deleting a large genomic region is direct and little or no residual transcript is left behind, the drawback with using this method is that it could generate phenotypes contributed by underlying DNA regulatory elements rather than due to the loss of the lncRNA itself, as discussed above. Neighbouring gene expression could be affected when specific domains or promoters embedded in the lncRNA locus are removed (Mancini-Dinardo *et al.,* 2006; Mohammad *et al.,* 2010; Sigova *et al.,* 2013). Compared to the former method, the insertion of a premature termination sequence in the gene body is less disruptive and produces a reduced lncRNA expression. It should be noted that the efficiency of termination cassettes is dependent on the site of insertion and does not prevent the act of lncRNA transcription (Latos *et al.,* 2012). Residual RNA expression can still be detected and may contribute to a different phenotype obtained. In addition, there is also a risk of modification to the spacing of underlying DNA elements which could have an effect on neighbouring gene expression.

The preceding experimental methodologies have demonstrated that there are many challenges when studying lncRNA function in vivo and it is more complex and difficult than that of protein-coding genes. Manipulation of lncRNA expression requires a good

understanding of the lncRNA genomic locus itself as well as its neighbouring genes. The absence of an observable phenotype does not necessarily equate to redundancy as shown in the examples above. In addition, the latent phenotypes displayed by lncRNAs such as *MALAT1* and *C130071C03Rik/LINC00461* highlight the importance of gene-environment interactions (Arun *et al.,* 2016; Kim *et al.,* 2018; Wang *et al.,* 2019).

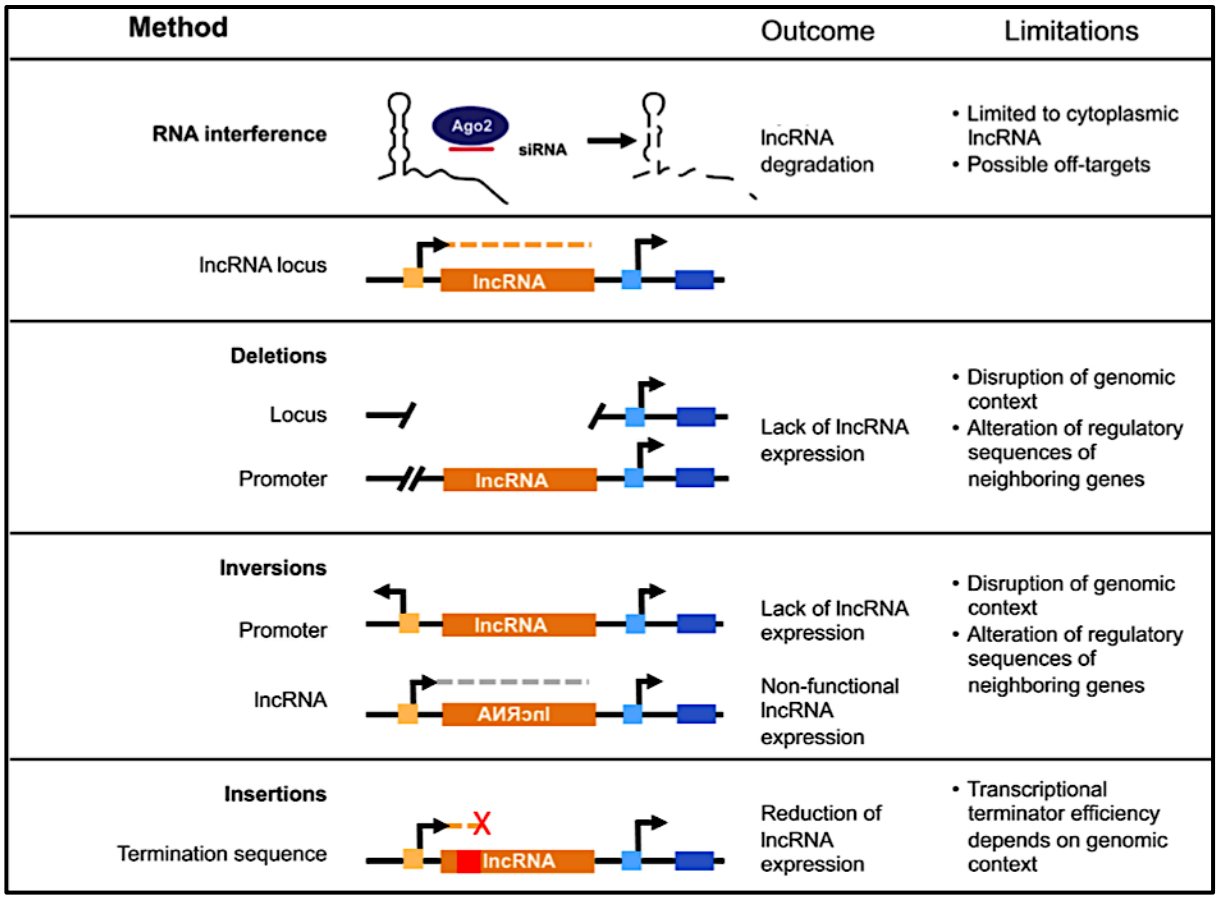


Figure 11: A schematic of genetic strategies to evaluate lncRNA functions in vivo.
The lncRNA locus is indicated in orange, neighbouring protein-coding gene in dark blue, transcription start sites are indicated by the direction of the arrows, the process of transcription by orange and grey dotted lines. RNAi is carried out using sh/siRNA sequences that bind to Argonaute proteins (*Ago2*, purple oval) within the cell and target lncRNA transcripts that are found in the cytoplasm for degradation. The deletion of the full-length lncRNA locus or its transcriptional start site sequences can be achieved using homologous recombination. This method can be combined with a reporter gene to create lncRNA null mutants and at the same time reveal in vivo expression patterns. The expression of the lncRNA of interest can be suppressed through the inversion of the lncRNA itself or its promoter sequence. Lastly, the insertion of a strong transcriptional stop signal, such as SV40 polyadenylation (polyA) signal or the chicken β-globin polyA signal, close to the start of the transcript unit allows early termination of the transcription. However, this method does not prevent the act of transcription and the ability to generate loss-of-function mutation is highly dependent on the genomic site.
[Image modified from Leone and Santoro, 2016]

1.11 Aim of the thesis

RNAs are not simply intermediate products of gene expression but play pivotal roles in many biological processes. The transcription machinery is pervasive throughout the genome of various organisms, generating a plethora of RNA transcripts. Non-coding RNAs have been shown to be key regulators of a range of biological processes relevant to development and disease. An increasing number of studies have described the involvement of lncRNAs in numerous processes such as epigenetic, transcriptional and post-transcriptional regulation. Little is known about the function of lncRNAs in *Drosophila* and how these lncRNAs are classified. The fruit fly was chosen as a model to study the molecular phenotype of the lncRNAs due to its unmatched flexibility for genetic manipulation. The stages of embryonic development are very well characterized and make an excellent model for studies of the roles of lncRNAs.

One of the many challenges faced in lncRNA studies is to resolve the issue of functionality of lncRNAs during animal development. The first aim of the project was to investigate the organismal functions of a limited set of lncRNAs through targeted disruption of their genomic loci. My approach was to first curate a list of conserved lncRNAs that show expression during *Drosophila* embryogenesis as observed from high-throughput sequencing data generated using the RAMPAGE (RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression) method. In order to systematically screen for functional lncRNAs, this study generated a catalogue of lncRNA mutants, using state-of-the-art genetic engineering technologies such as CRISPR/Cas9. A phenotypic analysis was subsequently performed to investigate the effects of lncRNA loss on the development and survival of *Drosophila* (chapter 3). LncRNA mutants that display overt phenotypes were further characterised in the following chapter.

The second part of the thesis is focused on the characterisation of one of the lncRNA mutants during embryogenesis (chapter 4). In this chapter, I made use of single-molecule fluorescent in situ hybridization (smFISH) to determine the cellular and intracellular localization of this lncRNA. This is followed by comparative RNA-Seq experiments between lncRNA mutants and wild-type flies to understand the specific targets of this lncRNA for further systematic studies. LncRNA-9 is required at stage 16/17 as the GO analysis suggests that it affects the expression of downstream genes in neuronal pathway associated with chemical synaptic transmission.

Chapter 2 |

Materials and methods

2.1 *Drosophila melanogaster* work

2.1.1 *Drosophila* Stocks

A list of *Drosophila melanogaster* strains used for genetic crosses or transgenesis in this study are listed in Tables 1, 2.1 and 2.2. Most fly stocks were obtained from the Vienna *Drosophila* Resource Center (VDRC) or the Bloomington *Drosophila* stock centre (Indiana University, Bloomington, USA). The Cas9 strains used were provided by the fly facility at the Department of Genetics, University of Cambridge.

2.1.2 *Drosophila* husbandry

All fly strains were maintained in standard fly food (cornmeal) that was purchased from the Department of Genetics, University of Cambridge. Egg production was stimulated through yeast (Red Star Active Dry Yeast) supplementation. Fly stocks were kept at either 18 ºC or 25 ºC at a 12 hr:12 hr light: dark cycle with 65 % - 70 % humidity.

2.1.3 Fly crosses

Males and virgin female flies were collected every 6 hrs after eclosion and were kept in separate vials. Depending on the nature of the cross, each vial generally contained around 10-20 flies and each bottle contained around 150-200 flies. For the crosses to be maintained for a longer period, the flies were flipped into a new vial with food and fresh yeast every 4 days. All genetic crosses were maintained at 25 ºC unless otherwise indicated.

2.1.4 Embryo microinjections and transgenesis

Microinjection into *Drosophila* embryos was carried out according to standard protocol by the University of Cambridge Department of Genetics Fly Facility. All plasmids used were purified using QIAGEN's Midi Plasmid purification kit (QIAGEN) and eluted in nuclease-free water (Thermofisher Scientific). The concentration of each plasmid was measured using a NanoDrop Spectrophotometer (Thermofisher Scientific) and the DNA samples had an OD260/280 ratio of 1.80-1.90. To generate transgenic lncRNA knockout lines, plasmid pools containing the donor construct and dual gRNA expression vector for each lncRNA were combined at a ratio of 2:1 respectively (Table 3 and 4). The plasmids were injected into TH_attp2 or vas-Cas9.RFP- (Bloomington Stock BL #55821) embryos and the RFP+ marker encoded by the donor plasmid was used to identify transgenic offspring. The offspring were subsequently balanced with respective balancer lines based on the genomic location of the lncRNA. To test whether the cassette was inserted into the appropriate locus of interest, the region flanking the deletion was amplified by PCR and sequenced for verification. For the flies with the SV40 polyA insertion, PCR was used to validate the insertion.

2.1.5 *Drosophila* embryo staging and collection

Agar plates containing apple juice hardened with agar were produced by the Department of Genetics, University of Cambridge and were used for embryo collection. To obtain developmentally staged embryos, ~200-300 adult flies (~4-5 days old) were transferred to a large embryo population cage (Flystuff #59-101) featuring a 90 mm apple juice agar plate containing ample amount of yeast paste. Prior to the actual collection, the flies were accustomed to the cage and transferred to new cages every day with fresh apple juice agar plates containing yeast paste. The apple juice plates were changed twice a day for two days. Embryo collections were performed after 2 x 1 hr pre-lays in the morning. Depending on the developmental stage required, the agar plates containing the embryos were incubated in the same incubator to allow embryogenesis to continue (0 – 24 hrs). The embryos were harvested using a paintbrush and rinsed in a 40 µM EASYstrainer (Greiner Bo-One) with deionised water. For the control, wildtype fly embryos were collected from a different cage from the lncRNA mutants. Each of the lncRNA mutant was crossed to a fluorescence balancer which allows stable populations of mutants to be maintained as balanced heterozygotes. The balancer chromosomes expressing the yellow fluorescent protein (YFP) allowed easy

identification, scoring and collection of live embryos with specific genotypes (Le *et al.,* 2006). The YFP signal can be detected from heterozygotes during embryonic development and this in turn allows us to differentiate the homozygous null mutants from the heterozygous siblings.

2.2 Molecular Biology Work

2.2.1 Polymerase Chain Reaction (PCR)

All PCR reactions were performed using KOD Hot start DNA Polymerase (Merck Millipore). A typical PCR reaction mix consist of 5 µL 10X buffer, 5 µL dNTP solution (2 mM), 5 µL DMSO, 3 µL primer mix (10 µM), 3 µL MgSO4 (25 µM), 1 µL KOD Hot Start DNA polymerase, 1-2 µL template DNA (1-100 ng/µL), and topped up to 50 µL with nuclease-free water (Thermofisher Scientific). The amount of DNA that was used for each PCR reaction varied depending on the template: ~10 ng of plasmid DNA or ~100 ng of genomic DNA was used for each PCR reaction. PCR reaction conditions were according to the manufacturer's instructions (Table 5). PCR products were analysed using a 1-2 % agarose gel (Invitrogen).

2.2.2 Primer design

Primers were designed using the Primer3web (version 4.0.0) online software. Standard DNA oligos were ordered and produced by Integrated DNA Technologies (https://eu.idtdna.com/site) or Sigma Aldrich (https://www.sigmaaldrich.com/life-science/custom-oligos.html). The genomic sequences used for primer design was obtained from the University of California, Santa Cruz (UCSC) Genome Browser (BDGP Release 5/dm3 assembly or BDGP Release 6 + ISO1 MT/dm6 assembly). The oligonucleotides were delivered as a lyophilised pellet and resuspended resuspended in nuclease-free water (Thermofisher Scientific) to make a stock solution of 100 µM. Oligonucleotides were stored at -20 ºC. A full list of oligonucleotides can be found in Tables 3, 4, 6 and 7.

2.2.3 High-quality genomic DNA extraction for molecular cloning

Genomic DNA was extracted from *Drosophila melanogaster* using the ZYMO (ZR) Research Tissue and Insect DNA Midiprep kit (Zymo Research). Approximately 300 flies

were snap frozen with liquid nitrogen and ground into a fine powder using a sterile mortar and pestle. The powder was transferred to a 15 mL falcon tube and mixed with 6 mL of Genomic lysis buffer containing 0.5% ß-mercaptoethanol. The suspension was mixed gently and transferred to a ZR Bashingbead Lysis/filtration tube. The suspension was manually shaken by hand for 3 mins before it was centrifuged at 4,000 × g for 5 mins. The lysate was filtered using a Zymo-Spin V-E column/Zymo-Midi Filter assembly fixed on a QIAvac vacuum manifold (QIAGEN) and the column containing the genomic DNA was transferred to a 2 mL collection tube and spun at 10,000 × g for 1 min. The flow through was discarded and 300 µL of DNA Pre-Wash Buffer was added to the column and spun for 1 min at 10,000 × g. The column was washed with 400 µL of DNA wash buffer and centrifuged twice at 10,000 × g for 1 min before it was transferred to a sterile 1.5 mL tube. Genomic DNA was eluted from the column by adding 150 mL of nuclease-free water (Thermofisher Scientific) and centrifuged for 1 min at 10,000 × g. DNA quantity and quality was assessed using a NanoDrop Spectrophotometer (Thermofisher Scientific). DNA was stored at -20 °C until further use.

2.2.4 Preparation of genomic DNA for genotyping

A separate protocol was used for fly genotyping. Single fly genomic DNA extraction was performed according to a protocol obtained from Dr. Yawen Chen in Dr. Stephen Cohen's Lab (IMCB, Astar Singapore). In brief, a single fly was placed in a 1.5 mL Eppendorf tube and frozen at -20 °C for at least 5 min or until processing. 50 µL of chilled Squishing buffer (10mM Tris-Cl pH 8.0, 1 mM EDTA, 25mM NaCl) was added to the Eppendorf tube with 1 µL of proteinase K (Thermofisher Scientific). The fly was ground against the tube wall using a nuclease-free pestle (VWR) and subsequently placed on a heat block for 45 min at 37 °C to allow protein digestion by the proteinase K. This was followed by heat-inactivation of the enzyme at 95 °C for 5. The sample was centrifuged at 14,000 × g for 10 min and the supernatant containing the genomic DNA was transferred to a fresh Eppendorf tube and stored at 4 °C for up to one week.

2.2.5 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by their sizes. Samples were mixed with 6 x purple loading dye (NEB) and the sizes of the fragments were compared to a Quick-Load® 100bp or 1 kb DNA Ladder (NEB). Agarose gels were cast with an appropriate amount (~0.5 µg/mL) of Ethidium Bromide solution (Invitrogen). The samples

were loaded into the gel and separated in 1x TAE buffer at 10V per cm gel length. DNA fragments were visualized by UV-light using a Bio-doc-it imaging system (UVP).

2.2.6 Purification of DNA with QIAquick gel extraction kit

DNA fragments from agarose gel slices were purified using the Qiaquick gel extraction kit (QIAGEN) according to the manufacturer's protocol. All centrifugation steps were performed at 17,900 x g for 1 min. In brief, gel slices were obtained under Visi-blue Trans-illuminator (UVP) and mixed with 3 volumes of Buffer QG per volume of gel. The samples were incubated at 50 °C for 10 min to allow for gel dissolving and then loaded onto a QIAquick column. The column was washed once with 500 µL of Buffer QG. The column was washed four times with 750 µL of Buffer PE. Residual ethanol was removed by an additional minute of centrifugation and the columns were transferred to a fresh 1.5 mL microcentrifuge tube. DNA was eluted using 30 µL of nuclease-free water (Thermofisher Scientific).

2.2.7 Purification of DNA with MinElute PCR purification kit

All centrifugation steps were performed at 17,900 x g for 1 min. 5 volumes of Buffer PB were added to 1 volume of PCR reaction and 10 µL of 3M sodium acetate at pH 4.8 (prepared by the media kitchen at the institute) was subsequently added. The sample was applied to a QIAquick column and centrifuged. The flow-through was discarded and the column washed with 750 µl of Buffer PE and then centrifuged for 1 min. The washing step was repeated three times. The flow-through was discarded and the column centrifuged again to remove trace ethanol. DNA was eluted by addition of 30-50 µl nuclease-free water (Thermofisher Scientific), followed by centrifugation for 1 min.

2.2.8 Nucleic acids quantification

The purity and concentration of DNA and RNA samples were assessed using a Nanodrop 1000 Spectrophotometer (Thermofisher Scientific). Purity of each sample was determined by the ratio of OD260:OD280. The acceptable readings for DNA samples was >1.8 and >2.0 for RNA samples. For a more accurate measurement of the concentration, the Qubit 3.0 Fluorometer (Thermofisher Scientific) was used for preparing libraries.

2.2.9 Sanger Sequencing

Sanger sequencing was carried out by GATC/Eurofins Genomics. Sequencing primers were designed according to the company's recommendations and only the template DNA and the sequencing primer were sent to the sequencing facility. The results ('*.seq*' and '*.ab1*' files) were analysed using an online software (www.benchling.com).

2.2.10 Quantification of expression levels by qRT-PCR

Quantitative PCR (qPCR) experiments were performed on a QuantStudio Real-Time PCR Light Cycler (Thermo Fisher Scientific). RNA levels were quantified using the ΔΔCT method (Livak and Schmittgen, 2001), normalized to Act5C and fold changes were calculated relative to the indicated controls. The primers used for qPCR can be found in Table 7.

2.3 Molecular Cloning

2.3.1 Digestion with restriction enzymes

Restriction enzymes were purchased from New England Biolabs (NEB). The NEB Double Digest Finder tool (http://www.neb.com/nebecomm/DoubleDigestCalculator.asp) was used to identify suitable buffers when two restriction enzymes were required. Restriction digests were carried out for 2-4 hours at the optimal reaction temperature (usually 37 °C). Each restriction enzyme reaction mix consists of 5 µL of 10x reaction buffer, 1 µL of DNA template (~1µg/µl), 1 µl of each restriction enzyme and topped up to 50 µL with nuclease-free water (Thermofisher Scientific). Plasmid digests were subsequently incubated with Alkaline Phosphatase, Calf Intestinal (CIP, NEB) for 0.5 hrs at 37 °C prior to separation on an agarose gel.

2.3.2 Construction of dual gRNAs targeting vectors

Dual guide RNAs (gRNA) were assembled and cloned into pCDF4 (Addgene #49411) (Port *et al.,* 2014). A pair of target specific sequence (sgRNA) were designed using the online CRISPR design tool (http://www.flyrnai.org/crispr/) for each target site. Only guides without off-targets were chosen. Using pCDF4 as the template, PCR products corresponding to each dual guide RNA were generated with KOD Hot Start DNA polymerase (Merck

Millipore) and purified with MinElute PCR purification kit (QIAGEN). The pCDF4 vector was digested with BbsI-HF (NEB) restriction enzyme and dephosphorylated with CIP (NEB) before gel purification. Cloning procedures were performed using 2x Gibson/HiFi Assembly Master Mix (NEB). The ligated product was transformed into One Shot TOP10 Chemically Competent bacteria (Invitrogen). To identify correct plasmids, the insert was screened by PCR using primers pCDF4_fwd and pCFD4_rev, and positive plasmids were subsequently verified by Sanger sequencing (GATC/Eurofin). Bacteria culture containing plasmids with the desired insert were used as starter culture for large-scale plasmid DNA isolation using QIAGEN Plasmid Midi/Maxiprep kit (QIAGEN).

2.3.3 Donor Vector Construct for HDR at each locus

CRISPR/Cas9 combined with homology directed repair (HDR) templates were used to create specific mutations or insertions. The donor vector contained a 3xP3-driven RFP marker to allow detection of a successful integration event. The pBluescript II KS (+/–) vector backbone and a plasmid containing the *3xP3-RFP-αtub-3'UTR* sequence were gifts from Drs. Fillip Port & Simon Bullock (LMB Cambridge). A pair of homology arms (~1-1.2 kb each) flanking the predicted double-strand break was generated from fly genomic DNA by PCR. The pBluescript II KS (+/–) vector, the backbone for the construct, was digested with HindIII-HF (NEB) and NotI-HF (NEB) restriction enzymes and dephosphorylated with CIP (NEB). A pair of PCR primers (*3xP3-RFP_fwd* and *3xP3-RFP_rev)* was used to amplify the *3xP3-RFP-αtub-3'UTR* fragment from the template plasmid. The PCR products were loaded onto a 1 % agarose gel and the DNA was extracted using a Qiaquick gel extraction kit (QIAGEN). NEBuilder HiFi DNA Assembly Master Mix (NEB) was used to assemble the four fragments into the final donor construct in a one-step reaction. A similar protocol was used to generate donor constructs with the transcriptional stop signal of *Simian vacuolating virus* 40 (SV40) instead of the *3xP3-RFP-αtub-3'UTR* sequence.

2.3.4 Cloning of single guide RNAs into the pCDF3 expression vector

A single gRNA (sgRNA) for the various lncRNAs was designed using the online CRISPR design tool (http://www.flyrnai.org/crispr/). The guide RNAs were assembled into a pCDF3 (Addgene #49410) expression vector as previously described (Port *et al.,* 2014). Briefly, the pCDF3 plasmid was cut with BbsI-HF (NEB) and purified using Qiaquick gel extraction kit (QIAGEN). A pair of oligonucleotides containing the sgRNA sequence was

phosphorylated using T4 PNK (NEB) and annealed. The phosphorylated, annealed fragment was ligated into the digested pCDF3 vector using T4 DNA ligase. The final construct was verified by restriction enzyme digestion and sequencing using pCDF3_seq_fwd and pCDF3_seq_rev primers.

2.3.5 Gibson/HiFi DNA assembly reaction

Gibson DNA assembly mix (NEB, 2x) and HiFi DNA assembly mix (NEB, 2x) were used to assemble multiple fragments into a vector. The reaction mixture consists of a mesophilic exonuclease, a thermophilic ligase and a high-fidelity polymerase, allowing for the rapid assembly of DNA fragments. A 2-3 factor excess of insert was used for every 50 ng of vector added. The assembly reaction was carried out at 50 °C for 1 hr and stored on ice before transformation into competent cells.

2.3.6 T4 PNK and T4 DNA ligation reaction

DNA ligation was used to assemble oligonucleotides containing a single sgRNA into a pCDF3 vector. A pair of oligonucleotides was annealed and phosphorylated in a reaction mix consisting of 1 µL of 100 µM oligonucleotide 1, 1 µL of 100 µM oligonucleotide 2, 1 µL of 10x T4 Ligation Buffer (NEB), 0.5 µL of T4 PNK (NEB) and 6.5 µL of nuclease-free water (Thermofisher Scientific). The samples were incubated in a thermocycler for 30 min at 37 °C, 95 °C for 5 min and then the thermocycler was set to cool down to 25 °C at 5 °C/min. The DNA fragment was ligated into a cut vector using T4 ligase. The ligation reaction consists of 1.5 µL 10 x T4 Ligation Buffer (NEB), 1 µL diluted fragment from the previous reaction (1:200 dilution), 1 µL of digested vector (50 ng/µL) , 1 µL of T4 DNA ligase (NEB) and topped up to 15 µL with nuclease-free water (Thermofisher Scientific). The ligation reaction was carried out at 16 °C overnight and transformed into One Shot TOP10 Chemically Competent bacteria (Invitrogen).

2.3.7 Transformation of commercial Endura electro-competent *E. coli*

Endura electro-competent cells were purchased from Lucigen and stored at -80 °C. The DNA assembly reaction was diluted 5 times and 1 µL was added to 20 µL of Endura Electro-competent cells. The mixture was pulsed with Bio-Rad Gene Pulser II (#165- 2105) in chilled 0.1 cm electroporation cuvettes (Bio-Rad). 980 µl of recovery media was added to

each cuvette and mixed with the pulsed cells. The cells were recovered at 37 °C for in a incubator for 1 hr with shaking at 225 rpm. 100 µL of reaction mix from each transformation was spread on a pre-warmed Lysogeny broth agar plate with appropriate antibiotics using plating pearls (Merck Millipore). The plates were incubated over night at 37 °C.

2.3.8 Transformation of commercial Chemically Competent bacteria

One Shot TOP10 Chemically Competent cells were purchased from Invitrogen and stored at -80 °C. The DNA ligation mixture was diluted 5 times and 1 µL was added to 25 µL of chemically competent cells. The vial was incubated on ice for 30 min, heat shocked at 42 °C for 30 sec and incubated again on ice for 2 min. 200 µL of pre-warmed SOC media was added and placed in a incubator for 1 hr with shaking at 225 rpm to allow the cells to recover. The culture was diluted 1:100 and 100 µL of diluted culture were spread on LB plates (prepared by the media kitchen at the institute) containing appropriate amounts of antibiotic. The LB plates were incubated at 37 °C overnight.

2.3.9 Plasmid DNA purification with the QIAprep Spin Miniprep Kit

A single colony was inoculated in 5 mL LB medium containing the appropriate antibiotic and incubated in a shaker incubator overnight at 37 °C. 2 mL of culture were pelleted by centrifugation at 6,800 x g for 5 min. The supernatant was discarded and the cells were thoroughly resuspended in 250 µL Buffer P1. Cell lysis was induced by adding 250 µL Buffer P2. After incubation for 3 min, 350 µL Buffer P3 were added and the reaction mix was centrifuged at 17,900 x g for 15 min to pellet cell debris and proteins. The cleared lysate was transferred into a QIAprep spin column and centrifuged at 17,900 x g for 1 min. 500 µL Buffer PB were added, followed by centrifugation at 17,900 x g for 1 min. The column was washed twice with 750 µL Buffer PE with centrifugation at 17,900 x g for 1 min between each wash. In order to remove residual ethanol, the column was centrifuged for an additional 2 min at 17,900 x g before it was transferred to a sterile 1.5 mL microcentrifuge tube. 50 µL pre-heated water were applied to the column and following incubation for 2 min at room temperature, DNA was eluted by centrifugation at 17,900 x g for 1 min. A Nanodrop 1000 Spectrophotometer (Thermofisher Scientific) was used to assess DNA quantity and quality.

2.3.10 High yield plasmid isolation with the QIAGEN Plasmid Plus Midi Kit

Large amounts of plasmid DNA were isolated from *E.coli* using the QIAGEN Plasmid Plus Midi Kit (QIAGEN) and the protocol provided. In brief, 350 µL of bacteria culture from the miniprep culture was inoculated into 350 mL of fresh LB medium containing the appropriate antibiotic and the culture was incubated in a shaker incubator overnight at 37 °C. Bacteria were pelleted by centrifugation at 4,000 rpm for 20 min at 4 °C. The pellet was thoroughly resuspended with 4 mL of Buffer P1 and lysis was induced with the addition of 4 mL Buffer P2. The reaction mix was incubated for 3 min at room temperature. 4 mL Buffer P3 was added to stop the lysis reaction and the sample was centrifuged at 4,000 rpm for 5 min at 4°C. The clear lysate was filtered through a QIAfilter Cartridge and 2 mL of Buffer BB was mixed with the filtered lysate. The mixture was transferred to a QIAGEN Plasmid Plus spin column on the QIAvac 24 Plus. The column was subjected to ~300 mbar vacuum until all the liquid was removed. The column was first washed with 700 µL of Buffer ETR followed by two washes with Buffer PE. Residual ethanol was removed by centrifugation at 10,000 x g for 2 min. 200 µL preheated water was applied to the column and after incubation for 2 min at room temperature, DNA elution was performed in a clean 1.5 mL Eppendorf tube by centrifuging at 10,000 x g for 1 min. The concentration and purity of the plasmid was assessed using Nanodrop 1000 Spectrophotometer (Thermofisher Scientific).

2.4 RNA-related experiments

2.4.1 Tissue lysis for RNA extraction

For each replicate, 100-200 embryos were collected and placed in a 1.5 mL Eppendorf tube containing 1 mL of chilled PBS. The embryos were centrifuged at 5,000 × g for 5 min at 4 ºC and washed with 1 mL of chilled PBS for two additional times. The tissue was homogenized in 50 µL of TRIzol Reagent (Invitrogen) for 1 min using a disposable RNase-free polypropylene pellet pestle and pestle motor mixer. An additional 100 µL of TRIzol Reagent was added to the lysate and, after grinding the sample by hand for 1 min, 850 µL of TRIzol Reagent were added and the tube was incubated at room temperature for 5 min before storing them at -80 ºC until further use.

2.4.2 RNA extraction and purification

Prior to all RNA work, the work area and pipettes were cleaned with RNase Zap (Sigma) to reduce the risk of RNase contamination. The samples were thawed on ice and centrifuged for 1 min at 12,000 × g before transferring them to a pre-spun 5PRIME Phase Lock Gel-Heavy tube (Quantabio). The lysate was incubated with the gel for 5 min at room temperature before adding 200 µL of chloroform to the tube and shaken vigorously for 30 sec. The samples were incubated for 3 min at room temperature and centrifuged at 12,000 × g for 15 min at 4 ºC. The upper phase was transferred to a new RNase-free tube containing an equal amount of RNase-free ethanol. The RNA was subsequently purified using the RNeasy Mini kit (QIAGEN) according to the manufacturer's protocol. Briefly, the sample was loaded onto a RNeasy column and spun at 10,000 × g for 30 sec. Buffer RW1 was added to the RNeasy spin column and centrifuged for 30 s at 10,000 × g. DNase I pre-mix containing 70 µL of Buffer RDD and 10 µL DNase I stock solution (QIAGEN) was added to the centre of the membrane and incubated for 15 min at room temperature. The spin column was loaded with Buffer RW1 and centrifuged for 30 s at 10,000 × g. Buffer RPE containing ethanol was used to wash the membrane and centrifuged at 10,000 × g for 1 min for two times. The column was transferred to a new 2.0 mL collection tube and spun at 10,000 × g for 2 min to remove excess ethanol. To elute the RNA, the spin column was transferred to a new 1.5 mL Eppendorf tube and 30 µL of nuclease-free water (Thermofisher Scientific) were added to the membrane to elute the RNA. Quantification of the RNA sample was performed using Qubit RNA BR (Broad-Range) Assay Kit (Thermofisher Scientific) and Qubit 3.0 Fluorometer (Thermofisher Scientific). The quality of the total RNA was analysed on the Agilent 4200 TapeStation system using the RNA ScreenTape assay (Agilent).

2.4.3 RNA sequencing (RNA-seq) library production

The RNA-seq libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep kit for Illumina (NEB) and NEBnext Poly(A) mRNA magnetic isolation module (NEB) according to manufacturer's protocol. Embryos from the control strain were collected at hourly intervals after egg laying (AEL) from 4 to 24 hrs. The mutant fly strain for the lncRNA mutant strain was crossed to a fluorescence balancer which is an important element for our experiment. The homozygous null mutant does not carry the balancer and will not express the fluorescence protein. On the other hand, a fluorescence signal can be detected in the heterozygous flies due to the expression of the YFP protein on one of the the balancer

chromosome. The mutant embryos were collected from 10 to 24 hrs by manually sorting these from the heterozygote YFP+ embryos (the balancer produces YFP protein in embryos). For each timepoint, the samples were collected in replicates of two. Total RNA was extracted from these two groups of embryos using the RNeasy Mini kit (QIAGEN) with DNase I (section 2.3.13) and kept at -80 ºC. For the library preparation, all the reactions were performed on ice unless otherwise stated. The Agencourt RNAClean XP system (Beckman Coulter) was used instead of the NEBNext Sample Purification Beads. The first part of the library preparation was aimed to enrich for polyA+ mRNAs and convert them to cDNA products. This means that any putative non-polyA targets will be missed in this RNAseq. Using 0.5 µg of total RNA as starting material, mRNA transcripts were isolated using oligo-dT-attached magnetic beads and fragmented to ~200 bp RNA inserts using divalent cations in the NEBNext First Strand Synthesis Reaction Buffer (5X) at 94 ºC for 15 min. The first strand cDNA synthesis reaction mix also contains NEBNext random hexamer primers and M-MuLV Reverse Transcriptase (RNase H-) to assemble the first stand cDNA. In the second cDNA synthesis reaction, NEBNext second strand synthesis reaction buffer containing dUTP, DNA Polymerase I and RNase H was added to the first strand synthesis product. The cDNA was purified and stored at -20 ºC until the next step. The DNA fragments with overhangs undergo a process of blunting followed by DNA dA-tailing so as to prepare them for ligation with NEBNext adaptors. The DNA fragments were then purified using the AMPure XP system and incubated with the USER Enzyme (NEB) supplied at 37 °C for 15 min followed by 95 °C for 5 min. To enrich for adaptor ligated DNA products, PCR was performed with Universal PCR primers, Index (X) Primer and NEBNext Ultra II Q5 Master Mix that contains Q5 High Fidelity DNA Polymerase. The PCR reaction was carried out according to the manufacturer's protocol and the PCR products were purified with AMPure XP system. The library was eluted in 20 µL of 0.1X TE (NEB) and stored at -20 °C. The library quality was assessed on the Agilent 4200 TapeStation system using the DNA ScreenTape assay (Agilent). The samples were quantified using the KAPA Library Quantification kit (KAPA Biosystems) and the pooled library was sequenced on a Hiseq 4000 at a mean depth of 30 million reads and 50bp single-end reads per sample. The RNA-seq run was performed by the CRUK Cambridge Institute genomics core facility.

2.5 Test of survival to adulthood

Heterozygous mutant stocks for each lncRNA were assessed for their viability. Before the assay was carried out, males and virgin females were collected in separate vials three days

earlier (section 2.1.3). After setting up a cross with heterozygous parents, the vials were kept at 25 ºC and the parents were transferred to a new vial after 24 hr for 2 more times. The progeny from each of the crosses was allowed to develop further and hatching adult flies from each of the three vials were counted after 14 days.

2.6 Embryo, Larval and Pupal Survival test

Wildtype, heterozygous and homozygous embryos were collected according to the method stated in section 2.1.5. Homozygous mutant progenies were identified by the absence of fluorescence expression from the balancer. The flies used as the control was collected from a different cage containing wildtype flies. A 4-hr collection window was used to collect large quantities of embryos. Apple juice plates containing the embryos were placed in the 25 ºC incubator and allowed to develop for another 12 hrs (12 hr -16 hr). For each genotype, embryos were transferred to 6 x 48-well plates containing cornmeal in each well. The plates were placed in the incubator for another 24 hrs and the number of first instar larvae that hatched were counted and transferred to a standard fly vial containing cornmeal. To determine the number of larvae that complete larval development and undergo pupation, the vial was kept in the 25 ºC incubator for another eight days and the number of larvae that became pupae were counted. To determine the number of pupae that completed pupation and hatched into adult flies, the vials were incubated for another 8 days and the emerging adults flies were counted.

2.7 Test of Fertility

The test of fertility (Yawen *et al.,* 2014) was used to differentiate the homozygous mutants into 3 categories: sterile (no progeny), low fertility (<10 larvae/vial) or fertile (≥10 larvae/vial). Using the observations from the viability test, the fertility test was only performed for lncRNAs that were viable as homozygous mutants. For each cross, a single fly mating strategy was used. In each vial, a single homozygous mutant male was mated with a single w^{1118} virgin female and vice versa. 10 crosses (5 mutant males and 5 mutant females) were set up for each lncRNA mutant. The parents were removed after 5 days and the number of larvae in each vial was counted.

2.8 Measurement of larval behaviour with a simple locomotion assay

A modified locomotion assay (Nichols *et al.,* 2012) was used to determine if the larvae showed any neuromuscular defects due to the lncRNA deletion. A 100 mm apple juice agar plate was used as the platform for the test. For each experiment, 50 first instar larvae were collected and arranged in one end of the plate that contained yeast paste on the other end. The number of larvae that moved toward the yeast was scored following 30 mins. The experiment was performed in five replicates. For each genotype, control larvae, heterozygous and homozygous lncRNA mutant larvae were used.

2.9 RNA fluorescence in situ hybridization (RNA-FISH) and immunohistochemistry

2.9.1 RNA-FISH probe synthesis

The reagents, nucleic acid probes and hairpin sequences used in this study were obtained from Molecular Instruments, Inc (www.molecularinstruments.org) and Advance Cell Diagnostics (ACD). The sequence of lncRNA-3 and lncRNA-9 were provided and probes were designed and synthesized by the companies.

2.9.2 Embryo dechorionation and fixation

Embryos of the appropriate stages were used for smFISH and IF (section 2.1.5). Embryos were placed in a 40 µM EASYstrainer (Greiner Bo-One, Austria) and rinsed gently with water to remove excess yeast. The embryos were then dechorionated by immersing the mesh in a solution of diluted bleach (50%) and shaken gently for 3 minutes. The embryos were washed with PBST (0.1 % Tween 20) to remove the remaining bleach solution and transferred to a 20 mL scintillation glass vial containing 8 mL of 4.5 % formaldehyde fixation solution (4.5 % formaldehyde, 0.5 x PBS, 25mM EGTA, 50% Heptane). The vial was rotated gently for 25-30 min to ensure that the embryos are exposed uniformly to the fixative solution. The lower phase containing the formaldehyde was removed subsequently and 8 mL of methanol were added to the vial and shaken vigorously by hand for 1 min. The vial was incubated for 2 min to allow the devitellinized embryos to sink to the bottom. The devitellinized embryos were transferred to a glass scintillation vial and rinsed three times with fresh methanol before being stored at -20 ºC until future use.

2.9.3 Florescence In situ hybridization

The spatio-temporal expression pattern for lncRNAs of interest were analysed using in situ HCR v3.0 probes from Molecular Instruments, Inc. The experiment was performed according to the protocol provided by the manufacturer with some modifications. In brief, the embryos from section 2.8.2 were transferred to a clean 1.5 mL Eppendorf tube and rinsed with ethanol before incubation in a xylene: ethanol mixture (1:1) for 2 hrs on a rotator. The supernatant was removed and the embryos were rinsed with ethanol followed by a series of methanol washes (100 %, 70 %, 50 %, 30 %). The embryos were washed with PBST and fixed in 4 % formaldehyde for 25 min. After fixation, the samples were washed with PBST and incubated with pre-hybridisation buffer for 30 min at 37 ºC. The pre-hybridisation buffer was removed and 100 µL of probe hybridisation buffer containing the probe set was added to the embryos and incubated for at least 16 hr at 37 ºC. After the hybridisation stage, the samples were washed with probe wash solution (pre-heated to 37 ºC) to remove excess probes followed with amplification buffer for 30 mins at room temperature. The amplification stage was performed in the dark and the samples were wrapped with aluminium foil to reduce exposure to light. The snap-cooled hairpins (95 ºC for 1 min 30 sec followed by 30 min at room temperature in the dark) were mixed with 100 µL of amplification buffer and added to the samples. The amplification stage took place in a dark room for 16 hr at room temperature and excess hairpins were washed away with 5 x SSCT at room temperature. The samples were mounted on slides using ProLong Diamond Antifade Mountant (Thermofisher Scientific) and visualized on a Leica Sp8 confocal microscope.

A separate set of probes from ACD was used to probe for lncRNA localization in sectioned embryos. Simultaneous detection of *Drosophila Act79d* and *Drosophila* lncRNA-9 was performed on embryo sections using RNAscope® 2.5 LS Duplex Reagent Kit (ACD). The fixed, dechorionated embryos stored in methanol were transferred to 70 % ethanol and embedded in paraffin. For each experiment, 4 µm sections of the embryos were collected. The sections were incubated at 60 °C for 1 hr and subsequently loaded onto a Bond RX instrument (Leica Biosystems). Prior to the treatment with Epitope Retrieval Solution 2 (Leica Biosystems), the slides were deparaffinised and rehydrated on board. The probe hybridization and signal amplification processes were carried out according to the manufacturer's instructions. The processed slides were mounted using VectaMount Permanent Mounting Medium (Vector Laboratories) and imaged on an Aperio AT2 (Leica Biosystems) at 40x magnification. The embedding, sectioning and staining of the embryos were performed

entirely by Bev Wilson and Julia Jones from the Histopathology/ISH Core Facility without any contributions from me.

2.9.4 Embryo Immunofluorescence

Primary antibodies were obtained from the Developmental Studies Hybridoma Bank (DSHB). Rat anti-Elav (1:100; 7E8A10, DSHB), mouse anti-Fas2 (1:100; 1D4, DSHB), and mouse anti-Repo (1:100; 8D12, DSHB) antibodies were used. Secondary antibodies were either goat anti-mouse or anti-rat conjugated to the Alexa 488 fluorophore (Thermofisher Scientific). For double labelling of whole-mount embryos, the samples prepared above were post-fixed with 4 % formaldehyde for 25 min and washed with PBST before antibody staining. The embryos were incbuated with 50% BSA on a rotator at room temperature for 1 hr. Primary antibody was added to the tube and incubated at 4 ºC for 12-16 hr. The embryos were subsequently incubated with secondary antibody overnight at 4 ºC. Hoechst 33342 (1:10,000) was used to stain the cell nuclei.

2.9.5 Image acquisition

Confocal imaging was carried out at the CRUK Cambridge Institute light microscopy core facility. Embryos were imaged using a Leica SP8 confocal microscope equipped with a 405 nm laser diode, white light laser, and hybrid detectors. The samples were imaged using 40× and 63 x glycerol objective with a ∼AU=1 pinhole. For each acquisition, 80-100 slices were imaged. Fluorescence and light microscope images of whole mount embryos, pupae and adult flies were acquired using a Leica fluorescence MZ10F fluorescence stereoscope fitted with a Leica DFC 7000T camera. The images have been obtained at 4x magnification and were processed using ImageJ and Fiji software.

2.10 Statistical analysis of phenotypic data

All statistical analysis was performed using Microsoft excel and prism GraphPad 8 unless otherwise stated. All data sets were compared to appropriate controls using Student's *t* test or chi-squared test ($\chi 2$). No statistical methods were used to predetermine the sample size used.

2.11 RAMPAGE library preparation and lncRNA selection

The RAMPAGE experiments, data analysis and lncRNA candidate selection was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me.

The full experimental design and analysis for the 5 *Drosophila* species was carried out by Dr. Philippe Batut and have been reported previously (Batut *et al.,* 2013; Batut and Gingeras, 2013; Batut *et al.,* 2017). In brief, 5 different *Drosophila* species were used in the study [*Drosophila ananassae* (#14024-0371.13)*, Drosophila erecta* (#14021-0224.01)*, Drosophila melanogaster* (#14021-0231.36), *Drosophila pseudoobscura* (#14011-0121.94) *and Drosophila simulans* (#14021-0251.195)]. The embryos were collected similarly to the protocol described in section 2.1.5 at hourly time points and aged appropriately. A total of 120 samples were collected by Dr. Batut and used for library production.

Briefly, extraction of total RNA was performed using a Beatbeater (Biospec) and 1.0 mm zirconia beads (Biospec) to lyse the embryos and then the RNAdvance Tissue kit (Agencourt) with DNaseI treatment for RNA extraction. TEX digest (Epicentre) was used in order to remove RNAs with 5'-monophosphate and the RNA was purified by RNAClean XP (Agencourt) according to the manufacturer's instructions. To ensure that the RNA used was of high quality, the Bioanalyzer RNA Nano chip (Agilent) was used prior to reverse transcription. A double-selection strategy allowed for precise identification of the transcriptional start site (TSS) along with sequencing of the transcript. A template-switching reverse-transcriptase was employed to enrich for 5'-complete cDNAs, while a cap-trapping technique was used to further select only the cDNAs retrotranscribed from capped RNA. Linker sequences were affixed to the ends of 5′-complete cDNAs and capped RNA molecules were biotinylated. Streptavidin-coated magnetic beads (PureBiotech LLC) were used to pull-down the biotinylated caps on the RNA along with the associated 5′-complete cDNAs. The purified libraries were subsequently subjected to PCR amplification followed by a size selection step. During the reverse transcription step, for every time series, a distinctive sequence barcode was incorporated to tag each sample. A single library was generated from pools of samples from the same time point. DNA High Sensitivity Bioanalyzer chip (Agilent) was used to check for the quantity and quality of the libraries produced. Paired-end sequencing was performed for each library using an individual lane on the Illumina HiSeq 2000 platform.

2.12 Bioinformatic Analysis - Quantification and statistical analysis

The bioinformatic analysis of RAMPAGE data in was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me.

2.12.1 RAMPAGE: primary data processing

The primary data processing in this section was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me. In brief, reference sequences and annotations used for analysis were acquired from Flybase (http://flybase.org) for *Drosophila ananassae* release 1.3, *Drosophila erecta* release 1.3, *Drosophila melanogaster* release 5.49, *Drosophila pseudoobscura* release 2.9 and *Drosophila simulans* release 1.4. Reads were trimmed by removing the sequences corresponding to the library identification barcode as well as the non-templated Gs resulting from the template switching and the reverse-transcription primer. Trimmed reads were subsequently mapped to the appropriate reference genome using STAR aligner (Dobin *et al.,* 2013). Filtering steps were applied to remove PCR duplicates and select only uniquely mapping reads. In order to identify TSSs, the density of 5' ends for each transcript as well as the coverage for the rest of the transcript was determined using the whole dataset. Then a peak caller algorithm was applied using a sliding-window approach to assess the statistical enrichment of 5' signal for each position in the genome considering a window surrounding that position. When multiple significant windows were found in adjacent to each other they were attributed to the same TSS cluster (TSC). TSC were then assigned to the best matching annotated gene and normalised for sequencing depth to give a precise measure of the promoter activity for each transcript (Batut *et al*., 2013; Batut and Gingeras, 2013). Peaks that were found to coincide with Flybase-annotated rDNA were removed (Batut *et al.,* 2017). The .sra files for each of the *Drosophila* species can be found in Table 8.

2.12.2 RAMPAGE: TSS clusters (TSCs) and conservation

This section of the work was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me. The TSC peaks identified in section 2.11.2 were subsequently evaluated for functional conservation. For each species, peaks passing certain criteria ($\geq$ 15 RAMPAGE tags, absent

in the heterochromatic regions or chromosome 4 in *Drosophila melanogaster*, or orthologous regions in other species) were selected for further analysis. To be able to compare different species, the genomic coordinates for each of the peak were transformed to coordinates in the multiple sequence alignment of all genomes (15-way MultiZ alignment from UCSC, http://hgdownload.soe.ucsc.edu/goldenPath/dm3/multiz15way). A list of parameters was then used to assess if each peak possesses unique syntenic alignment in all other species. From the middle of the peak, both ends of an 800 bp window had to map to the same strand of the same chromosome. Peaks with less than 50% of bases aligned (i.e., not in assembly gaps) and less than 25% of bases matching orthologous bases (not alignment gaps) were discarded. Only syntenic peaks were used for further analysis and, in order to compare the peaks in different species, the raw 5' signal for these peaks in each genome was also translated into multiple alignment coordinates. The RAMPAGE tags for each peak from each species were then counted to evaluate conservation across species. A peak was considered absent in one species if it had 100-fold lower signal than in the reference species. In addition, any peak that had <100 tags in the reference species was also filtered out if no signal was detected in a target species (Batut *et al.,* 2017).

2.12.3 RAMPAGE: Reconstruction of lncRNA transcripts and ORF analysis

This section of the work was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me. Focusing mainly on lncRNAs, De novo assembly of lncRNA transcripts was performed using Cufflinks (version 2.2.1) and Cuffmerge with default parameters on a recently published RNA-seq developmental time course (Graveley *et al.,* 2011). Each published dataset in the developmental series was ran independently using the program. For a transcript model to be identified as a RAMPAGE TSC, a boundary of 150 bp from the 5' end of the transcript was used. Models that did not match any TSC were filtered out and omitted from future analyses. In order to distinguish coding and non-coding transcripts, coding potential analysis was performed with phyloCSF on the annotation sets and the 15-way multiZ whole-genome alignments (Lin *et al.,* 2011). Positive phyloCSF scores were indicative of protein-coding genes whereas non-coding transcripts scored negatively. The coding capability of our lncRNA gene set was also measured using *CPAT* version 1.2.1 (Wang *et al.,* 2013). Following the developers' indications, a cut-off value of 0.39 was taken to assign evidence of coding potential by this method. If the value of a transcript was less than 0.39, it was considered to be a ncRNA. Another method used to exclude coding potential was Coding-Non-Coding-Index

(CNCI) (Sun *et al.,* 2013). Transcripts with coding potential greater than the cutoff score of 0 were filtered out, and those with a score of less than 0 were considered to be without coding potential.

2.12.4 RAMPAGE: Time series alignment

The time series alignment work was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me. In order to compare gene expression for the different species on a relative time scale, an alignment of time series data (or time warping) was applied. A Z-score transformed gene expression was computed for each species and for the entire time series and they were registered to one another according to the recommendations of the software documentation (Goltsev and Papatsenko, 2009). The gene expression profiles between species were paired using one-to-one orthology calls from Flybase (2012 release 2). In order to offset for the variances in annotation quality and peak calling between species, sets of datasets (*Drosophila melanogaster* and another species) were subjected to preliminary processing. Sometimes this brought to the identification of additional orthologous TSCs that were expressed (≥10 tags) but initially missed in one of the two species. In the event that a functionally conserved peak was attributed to an annotated gene in only one of the species pair, this inconsistency was corrected by assigning it the same gene in both species. A 5-fold up-sampling was applied to all of the time series and a 2-hr window size was used for data smoothing with the RZ-Smooth version 4.1 program. The UCSC liftOver tool (http://hgdownload.cse.ucsc.edu/....x86_64/liftOver) was used to align *Drosophila melanogaster* TSCs to all the other genomes. An alignment is considered to be good if it has ≥50% of bases aligned. Only the temporal expression profiles of these orthologous genomic positions were aligned. T-Warp version 3.2 with Pearson distance matrices (3 hr window) was used to optimised the global alignment paths between *Drosophila melanogaster* and the other datasets. The alignment of each series with *Drosophila* melanogaster as the reference was performed using M-Align version 2.8 and the data was subsequently smoothened for the final aligned series (1 hr window). The expression profiles of individual TSCs were registered to one another with M-Align, using the optimal alignment path computed for gene expression profiles. The 8th time point in the *Drosophila ananassae* dataset was omitted from further analysis because of poor data quality (Batut *et al.,* 2017).

2.12.5 RAMPAGE: Conservation analysis and assembly of TSC expression profiles

This section of the work was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab at Cold Spring Harbor Laboratory, USA without contributions from me. The conservation of individual expression profiles (TSCs or genes) across a clade was defined as the mean Pearson R^2 value for all pairwise comparisons of species within the clade. Functionally conserved TSCs in all five species (n = 240 and n = 3,824, respectively) were identified and were used for the selection of lncRNA candidates.

2.13 Bioinformatic Analysis - Quantification and statistical analysis

The mapping of the modENCODE data and analyses of the RNA-seq data were performed by colleagues at our in-house bioinformatics core facility team (Dr. Ashley Sawle and Dr. Abigail Edwards).

2.13.1 De novo transcript assembly and transcript mapping

(GEO Accession GSE36212) [Batut *et al.,* 2013] and 32 modENCODE RNA-seq libraries that were submitted by the Berkeley *Drosophila* Genome Project (BDGP) group were retrieved from the SRA database. These data were produced as part of the modENCODE project with high quality (http://www.modencode.org/). Before mapping, quality checking of raw reads (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) were performed using FASTQC version 0.11.5 (Table 9 and 10). Sample reads were aligned to the *Drosophila melanogaster* genome dm6 (corresponding to BDGP6) using tophat2 version 2.1.0 [Kim *et al.,* 2013] with options

```
tophat --GTF dme.dm6.gtf --bowtie1 --min-anchor 3 --num-threads 4 --
output <OutputBam> <Tophat_dm6_index> <FASTQ_Read1> <FASTQ_Read2>
```

The alignment QC was carried out using from Picard tools version 1.122 (https://broadinstitute.github.io/picard/command-line-overview.html). The tools CollectAlignmentSummaryMetrics and CollectRnaSeqMetrics were used with the following options, respectively:

```
java -jar picard.jar CollectAlignmentSummaryMetrics I=<InputBam>
O=<OutputFile> REFERENCE_SEQUENCE=<dm6 FASTA file>
VALIDATION_STRINGENCY=SILENT MAX_INSERT_SIZE=100000 ASSUME_SORTED=true
```

```
java -jar picard.jar CollectRnaSeqMetrics I=<InputBam> O=<OutputFile>
REF_FLAT=<ReferenceFile> STRAND_SPECIFICITY="None" ASSUME_SORTED="false"
VALIDATION_STRINGENCY=SILENT
```

Using the coordinates of the lncRNAs that were obtained using RAMPAGE, we converted them to dm6 coordinates using the UCSC lift-over tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). For each library, de novo transcripts were called using Trinity version 2.4.0 [Haas *et al.,* 2013], as follows:

```
Trinity --left <Read1.fq> --right <Read2.fq> --seqType fq --max_memory 8G
--SS_lib_type RF --no_version_check --output <OutputDirectory>
```

Well-defined genes were used as positive controls during the transcript assembly (*HERC2, pigeon, IA-2, CG10189, nkt, CG9706, luna, CG32181, Shrm*). The coding potential of the transcripts was assessed using the Coding Potential Calculator (CPC, score < 0, Kong *et al.,* 2007), Coding-Potential Assessment Tool (CPAT score < 0.403, Wang *et al.,* 2013) as well as Coding-Non-Coding Index (CNCI, score<0, Sun *et al.,* 2013).

2.13.2 Differential Gene expression analysis for lncRNA-9

The protocol used for the RNA extraction and purification is described in section 2.4.1 and 2.4.2. The embryos were first homogenized in TRIzol Reagent (Invitrogen) and the RNA was purified using the RNeasy Mini kit (QIAGEN) according to the manufacturer's protocol. A total of 56 RNA-seq libraries were first prepared using NEBnext Poly(A) mRNA magnetic isolation module (NEB) kit for polyA enrichment followed by NEB NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina kit according to the according to the manufacturer's protocol (and as described in section 2.4.3). The sequencing reaction was carried out on a Illumina HiSeq 4000 using single end (SE) 50 bp reads by the genomics core facility at the CRUK Cambridge Institute. For each replicate in the time series, all of the samples exceeded the recommended depth of 20 million reads for differential gene expression with at least 30 million reads per replicate obtained. The strand-specific, single-end raw sequencing reads were downloaded and the quality of these raw reads were verified using FASTQC version 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The reads were subsequently aligned to the *Drosophila melanogaster* reference genome with HISAT2 version 2.1.0 [Kim *et al.,* 2013] against dm6 using the following options:

```
Hisat2 -x <indexFILE> -U <inputFg> --summary-file <summaryFileName> -p 6
-rg-id Rg1 -rna-strandness R
```

To check for alignment quality, we used Picard tools version 2.60 (https://broadinstitute.github.io/picard/command-line-overview.html), with the following options for CollectRnaSeqMetrics and CollectAlignmentSummaryMetrics, respectively:

```
java –jar picard.jar CollectRnaSeqMetrics I=<InputBam> O=<OutputFile>
REF_FLAT=<ReferenceFile> STRAND_SPECIFICITY=”None” ASSUME_SORTED=”false”
VALIDATION_STRINGENCY=SILENT

java –jar picard.jar CollectAlignmentSummaryMetrics I=<InputBam>
O=<OutputFile> REFERENCE_SEQUENCE=<dm6 FASTA file>
VALIDATION_STRINGENCY=SILENT MAX_INSERT_SIZE=100000 ASSUME_SORTED=true
```

Alignment rates of the libraries to the *Drosophila* dm6 genome ranged from 84.5% to 96.7% (at least 95% in a good quality RNAseq library) and the intragenic alignment rates for the libraries were just slightly lower than the standard (95% or higher) at 93.9%.

Once the reads have been aligned, gene-specific read counts were performed with FeatureCounts from the Subread package v1.5.2 against the dm6 annotation [Liao *et al.,* 2014] and the predicted de novo transcripts from Trinity, with the following options:

```
 FeatureCountsBinary –s 2 --primary –C –T 4 –S –a <.saf > –o <outputFile>
<BamFile>
```

Reads overlapping with two or more features (regions of the genome where two or more genes overlap) were listed as ambiguous and not counted. The threshold for determining if a gene was genuinely detected was set at a minimum of five reads. There were 16,750 features (genes) in the annotation that were covered by at least one read in at least one sample. Coverage profile plots were generated to check for uniformity and to identify the presence of any 3’/5’ bias which may arise due to RNA degradation or other issues during library preparation. In the RNA sequencing libraries that were made, a uniform coverage plot was observed for all of the samples, indicating good quality RNA-Seq libraries. In order to further investigate correlation of counts between replicate, the distribution of correlation per gene count between sample pairs was examined. The raw counts were normalised to transcript per million (TPM) and the Pearson correlation coefficient for all possible sample pairs was calculated. From the analysis, lncRNA mutant samples collected from 22-23 hr were the worst correlated sample pair with a correlation score of 0.8558841. Differentially expressed transcripts/genes (DEGs) were predicted in R version 3.3.3 with the defaults of DESeq2 in a per time point analysis. During the differential analysis, default parameters were used with the ‘DESeq’ command. When determining if a gene was statistically significantly differentially expressed, we used a p-adjusted value (*p-adj*) cutoff of 0.05.

2.13.3 Gene ontology (GO) analysis

GO functional enrichment analysis is a commonly used tool to interpret functional roles of large-scale transcriptomics data (Ashburner et al., 2000; The Gene Ontology Consortium, 2018). The differentially expressed gene lists for each time point were subjected to enriched GO categorisation that characterises the relationship between genes and categorize them into broad categories such as molecular function, biological processes and cellular component. The Gene Ontology Consortium's online tool (http://www.geneontology.org/) was used to carry out the enrichment of our gene list (PANTHER version 15.0 Released 2020-02-14). During the analysis, an overrepresentation test was performed using the PANTHER Overrepresentation Test (Released 20200407) and all of genes in the database for *Drosophila melanogaster* (reference List). The gene names from our list were pasted into the GO tool and I have selected the species (*Drosophila melanogaster*) and specific ontology (GO-Slim molecular function, GO-Slim biological process and GO-Slim cellular component) for our enrichment analysis. At the end of each analysis, the various GO terms associated with the gene set were presented along with information on their degree of fold enrichment, a p-value, and a false discovery rate (FDR). When a GO term is over represented, it will be shown as a positive value and a negative value when it is underrepresented. The GO terms will be ranked from the highest to the lowest fold enrichment and only results with a FDR < 0.05 is displayed (Muruganujan et al., 2013).

Table 1: *Drosophila* stocks used for genetic crosses or transgenesis.

Control Stocks		
Name	**Source**	**Reference**
w[1118]	Gift from The University of Cambridge Department of Genetics Fly Facility	
OregonR	Gift from The University of Cambridge Department of Genetics Fly Facility	
Stocks with balancer chromosomes		
Df(1)m13, Gs2[m13]/FM7a, w[+]	The University of Cambridge Department of Genetics Fly Facility	
w[1118]/Dp(1;Y)y[+]; sna[Sco]/SM6a	The University of Cambridge Department of Genetics Fly Facility	
Pri[1] Dr[1]/TM3	The University of Cambridge Department of Genetics Fly Facility	
w[1118]; If/Cyo ; MKRS/TM6B, Tb[1]	Gift from The University of Cambridge Department of Genetics Fly Facility	
y1 w baz4 P{FRT(whs)}9-2/FM7c, P{Dfd-GMR-nvYFP}1, sn+*	[2]Stock ID: BL #23229	Tien *et al.*, 2006
w[*]; sna[Sco]/CyO, P{Dfd-GMR-nvYFP}2	[2]Stock ID: BL #23230	Tien *et al.*, 2006
w[*]; ry[506] Dr[1]/TM3, P{Dfd-GMR-nvYFP}3, Sb[1]	[2]Stock ID: BL #23231	Tien *et al.*, 2006
Cas9-expressing stocks		
y^1 M{vas-Cas9.RFP-}ZH-2A w^{1118}	The University of Cambridge Department of Genetics Fly Facility	
y[1] sc[1] v[1];; {y[+t7.7] v[+t1.8]=nanos-Cas9}attp2	[3]The University of Cambridge Department of Genetics Fly Facility	Port *et al.*, 2015
Deficiency lines		
w[1118]; Df(3L)BSC414/TM6C, Sb[1] cu[1]	[2]Stock ID: BL #24918	
w[1118]; Df(3L)BSC432/TM6C, Sb[1] cu[1]	[2]Stock ID: BL #24936	
w[1118]; Df(3L)ED4457, P{w[+mW.Scer\FRT.hs3]=3'.RS5+3.3'}ED4457/TM6C, cu[1] Sb[1]	[2]Stock ID: BL #9355	
Stock with microRNA gene mutation		
w; TI{TI}mir-276aKO/TM3, P{GAL4-twi.G}2.3, P{UAS-2xEGFP}AH2.3, Sb1 Ser1*	[2]Stock ID: BL #58906	Chen *et al.*, 2014
Other lines		
y1 v1 P{nos-phiC31\int.NLS}X; P{CaryP}attP40	[2]Stock ID: BL #25709	

[1]Flies obtained from Vienna Drosophila Resource Center (VDRC) (Vienna BioCenter, Vienna, AUSTRIA).
[2]Flies obtained from Bloomington Drosophila stock centre (Indiana University, Bloomington, USA).
[3]Fly was generated by Dr. Fillip Port and provided by the fly facility at the Department of Genetics, University of Cambridge.

Table 2.1: List of transgenic flies generated during this study.

Stocks containing 3xP3 transgene (KO lines)				
Name		gRNA plasmid ID	Donor plasmid ID	Mutation Type
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	2M	pES03	pES33	TSS-deletion
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	3M	pES03	pES33	TSS-deletion
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	6M	pES03	pES33	TSS-deletion
w; lncRNA-4 [RFP+]/CyO-YFP; +*	1M	pES04	pES34	TSS-deletion
w; lncRNA-4 [RFP+]/CyO-YFP; +*	3M	pES04	pES34	TSS-deletion
w; lncRNA-4 [RFP+]/CyO-YFP; +*	5M	pES04	pES34	TSS-deletion
lncRNA-5[RFP+]/Fm7a-YFP; +; +	1M	pES05	pES35	TSS-deletion
lncRNA-5[RFP+]/Fm7a-YFP; +; +	2M	pES05	pES35	TSS-deletion
lncRNA-5[RFP+]/Fm7a-YFP; +; +	3M	pES05	pES35	TSS-deletion
w; +; lncRNA-6 [RFP+]/Tm3-YFP*	2M	pES06	pES36	TSS-deletion
w; +; lncRNA-6 [RFP+]/Tm3-YFP*	4M	pES06	pES36	TSS-deletion
w; +; lncRNA-9 [RFP+]/Tm3-YFP*	1M	pES09	pES39	TSS-deletion
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	1M	pES10	pES40	TSS-deletion
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	3M	pES10	pES40	TSS-deletion
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	5M	pES10	pES40	TSS-deletion
w; lncRNA-11 [RFP+]/CyO-YFP; +*	1M	pES11	pES41	TSS-deletion
w; lncRNA-11 [RFP+]/CyO-YFP; +*	2M	pES11	pES41	TSS-deletion
w; lncRNA-11 [RFP+]/CyO-YFP; +*	3M	pES11	pES41	TSS-deletion
w; lncRNA-11 [RFP+]/CyO-YFP; +*	4M	pES11	pES41	TSS-deletion
w; lncRNA-12 [RFP+]/CyO-YFP; +*	1M	pES12	pES42	TSS-deletion
w; lncRNA-12 [RFP+]/CyO-YFP; +*	2M	pES12	pES42	TSS-deletion
w; lncRNA-12 [RFP+]/CyO-YFP; +*	3M	pES12	pES42	TSS-deletion
w; lncRNA-12 [RFP+]/CyO-YFP; +*	4M	pES12	pES42	TSS-deletion
lncRNA-13[RFP+]/Fm7a-YFP; +; +	1M	pES13	pES43	TSS-deletion
lncRNA-13[RFP+]/Fm7a-YFP; +; +	3M	pES13	pES43	TSS-deletion
lncRNA-13[RFP+]//Fm7a-YFP; +; +	5M	pES13	pES43	TSS-deletion
w; lncRNA-14 [RFP+]/CyO-YFP; +*	1M	pES14	pES44	TSS-deletion
w; lncRNA-14 [RFP+]/CyO-YFP; +*	2M	pES14	pES44	TSS-deletion
w; lncRNA-14 [RFP+]/CyO-YFP; +*	3M	pES14	pES44	TSS-deletion
w; lncRNA-14 [RFP+]/CyO-YFP; +*	4M	pES14	pES44	TSS-deletion
w; lncRNA-15 [RFP+]/CyO-YFP; +*	1M	pES15	pES45	TSS-deletion
w; lncRNA-15 [RFP+]/CyO-YFP; +*	2M	pES15	pES45	TSS-deletion
w; lncRNA-15 [RFP+]/CyO-YFP; +*	3M	pES15	pES45	TSS-deletion
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	2M	pES16	pES46	TSS-deletion
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	3M	pES16	pES46	TSS-deletion
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	4M	pES16	pES46	TSS-deletion
w; +; lncRNA-19 [RFP+]/Tm3-YFP*	1M	pES19	pES49	TSS-deletion
w; +; lncRNA-19 [RFP+]/Tm3-YFP*	2M	pES19	pES49	TSS-deletion

Table 2.2: List of transgenic flies generated during this study.

Stocks with SV40 insertion downstream of lncRNA TSS				
Name		gRNA plasmid ID	Donor plasmid ID	Mutation Type
w; +; lncRNA-9 [SV40]/Tm3-YFP*	2M	pES103a	pES107c	SV40 insertion
w; +; lncRNA-9 [SV40]/Tm3-YFP*	3M	pES103a	pES107c	SV40 insertion
Stocks expressing dual gRNA construct				
Name		gRNA plasmid ID		Mutation Type
y1 M{vas-int.DM}ZH-2A w; P{CaryP}attP40-white*	1M	pES_26		Insertion of dgRNA against white
Stocks expressing BAC construct				
Name		BAC plasmid ID	Mutation Type	Remarks
[1]*P{CH322-18B08}attP40/CyO*	2M	CHORI-CH322 -18B08	BAC insertion	Venken *et al.*, 2009
[1]*P{CH322-151D21}attP40/CyO*	3M	CHORI-CH322 -151D21	BAC insertion	
[1]*PBac{CH322-170N22}VK00001/CyO*	2M	CHORI-CH322 -170N22	BAC insertion	
[1]*PBac{CH322-122O7}VK00002/CyO*	1M	CHORI-CH322 -122O7	BAC insertion	
Other Stocks generated				
w[1118]; If/Cyo ; lncRNA3[RFP+]/TM6B, Tb[1]				
w[1118]; If/Cyo ; lncRNA9[RFP+]/TM6B, Tb[1]				

[1]Flies generated by BestGene.Inc, USA. All of the other flies were generated by the University of Cambridge Department of Genetics Fly Facility.

Table 3: List of gRNA plasmid and the primers used to generate the construct.

Dual gRNA plasmid list					
Candidates	**[1]sgRNA_1**		**[2]sgRNA_2**		**Dual-gRNA Plasmid #**
	Sequence	**Genomic Coordinates**	**Sequence**	**Genomic Coordinates**	
lncRNA-3	TATTAAAACGACTGGGGGGG	chr3L:10,325,557-10,325,579	ATGTGATATGTTGTGGGCTG	chr3L:10,327,299-10,327,321	pES_3
lncRNA-4	GTGTACTTACAGATTTCCAT	chr2R:10,263,671-10,263,693	GATGCGAAAGTATTCGATGG	chr2R:10,263,888-10,263,910	pES_4
lncRNA-5	AGGGTCTTTGTAGGTCTGCA	chrX:3,861,568-3,861,590	AGTAAATTTAATCAGAGACC	chrX:3,862,551-3,862,573	pES_5
lncRNA-6	AGGGACATACATCTAAGGAA	chr3R:16,772,794-16,772,816	GGGATTGAGCCAAGTTAAGG	chr3R:16,773,734-16,773,756	pES_6
lncRNA-9	TCGCTTGTCTGGTTTCGAACA	chr3L:17,269,326-17,269,346	AAGGCTCTCCATTAGCTTAA	chr3L:17,270,385-17,270,407	pES_9
lncRNA-10	CGATGTGTAACGTGAGAGCT	chr3L:13,112,379-13,112,401	GAAGTTTCCACTCATTTATC	chr3L:13,113,943-13,113,965	pES_10
lncRNA-11	CGGAAGGAGCCTGTATCCAT	chr2L:20,483,035-20,483,057	AAAATGTTAAAAATTTACTG	chr2L:20,485,049-20,485,071	pES_11
lncRNA-12	AAGCTACTAAACGATACAGA	chr2L:17,584,436-17,584,458	GCGAGGGTGAACCTGAACCC	chr2L:17,585,337-17,585,359	pES_12
lncRNA-13	GGCTGAAAACCCCTGGACCG	chrX:7,265,450-7,265,472	GTGGACGAGTGATTTCAATT	chrX:7,266,704-7,266,726	pES_13
lncRNA-14	GTGCATTATCGCGGATAATT	chr2R:9,549,757-9,549,779	GCAATTATCATCGTGCAGGT	chr2R:9,550,909-9,550,931	pES_14
lncRNA-15	ATGAAAAAAACCCGCCGACA	chr2R:13,362,387-13,362,409	GCCGAAGTTCATGCGTCCTT	chr2R:13,363,152-13,363,174	pES_15
lncRNA-16	AGTTACGGGTTACGGGAATT	chr3R:26,693,831-26,693,853	TTAAAATAACTGAAACCACT	chr3R:26,694,666-26,694,688	pES_16
lncRNA-19	CACCCTAACGGCAGCCTAAC	chr3R:16,847,383-16,847,405	TGTTTGTTGATAACACACTT	chr3R:16,849,367-16,849,389	pES_19
white	CGGGCCTCCCTCATAAAAAC	chrX:2,791,303-2,791,325	AATTGATGGCGTAAACCGCT	chrX:2,796,428-2,796,450	pES_26
Single gRNA plasmid list					
Candidates	**sgRNA_1**		**sgRNA_2**		**sgRNA plasmid #**
	Sequence	**Genomic Coordinates**	**Sequence**	**Genomic Coordinates**	
lncRNA-9	AAATAGAAGAAAGACTGCCA	chr3L:17,269,915-17,269,934	-	-	pES103a

[1]'TATATAGGAAAGATATCCGGGTGAACTTC' was added to the 5' end of the first gRNA and 'GTTTTAGAGCTAGAAATAGCAAG' was added to the 3' end.
[2] 'GTTTTAGAGCTAGAAATAGCAAGTTAAAAT' was added to the 5' end of the second gRNA 'GACGTTAAATTGAAAATAGGTC' was added to the 3' end.

Table 4: List of primers used for generating the donor construct.

Candidates	[1]Left homology arm primer 1	[2]Left homology arm primer 2	[3]Right homology arm primer 1	[4]Right homology arm primer 2	Plasmid #
lncRNA-3	TACGTTTCTGGGTGTGTCCAT	CCCCAGTCGTTTTAATATTATTTGCCT	CCCACAACATATCACATCCATATCC	CACTGATTTCGCCGATTTTTGTTG	pES_33
lncRNA-4	ACACTTACCAAATCCGAAATCATGA	GAAATCTGTAAGTACACAACTTGTGC	TCGAATACTTTCGCATGACGC	GAACTGAACTGAGGATGAGTGCGTG	pES_34
lncRNA-5	GGCTTTTGTGATGAATAGTTCGTTTAG	AGACCTACAAAGACCCTATGGC	CTCTGATTAAATTTACTTACAACTGCTGTT	CTTGGAGTTCTCCTACTCACAGT	pES_35
lncRNA-6	CGATAAAGGGTACGAGAAGAAGC	CTTAGATGTATGTCCCTTGGTCCA	TAACTTGGCTCAATCCCAGCT	GGGAATTGGAAAGGAGTGGAAC	pES_36
lncRNA-9	ACATCATTGTTGATGCTGCAATTAC	GACCCGAAAAAAACAGAAGAAAAATACG	AGCTAATGGAGAGCCTTATCTGTG	AAAGTGATGGTGGACACCAAAAG	pES_39
lncRNA-10	GAAGTTGCCTCCGTCTCAATC	TCTCACGTTACACATCGGAAAG	AAATGAGTGGAAACTTCGAGTGAA	TCAGTTTTCCATAGTGCTGCG	pES_40
lncRNA-11	TATGTAGGACTCAAAAAGTGTCAATCC	GATACAGGCTCCTTCCGTTTGG	TAAATTTTTAACATTTTCGTAGCCAGTAACG	CACTTTCTGTTGCAAATTGCGC	pES_41
lncRNA-12	ATTAATGGATGTCGTGGGTGTAGC	GTATCGTTTAGTAGCTTCCTTCTTTCG	TTCAGGTTCACCCTCGCTGTTT	TATTGGCAATATTTATGGCAGCCAT	pES_42
lncRNA-13	TTTTCAAGCAGGAGGCATACGA	TCCAGGGGTTTTCAGCTTTTTTG	TGAAATCACTCGTCCATTGTTTTTTC	ATCTTCGAGTTTCTTGAGATACCTAGC	pES_43
lncRNA-14	CTCCTTTGGTTACCAATTAGGCC	TATCCGCGATAATGCACAGCT	TGCACGATGATAATTGCATAAAATTACAG	AATCCGCACGGGAAGAAAACG	pES_44
lncRNA-15	TTAATTTAACGTCTTTATTGTAATTATAGAATTGCCA	CGGCGGGTTTTTTTCATGCTT	GTTCATGCGTCCTTCGGGTG	AGGCCTGTCTGCCTCTTTATTT	pES_45
lncRNA-16	AGAGGAGTTACTGCCTGAACC	TCCCGTAACCCGTAACTTGAG	GGTTTCAGTTATTTTAATAACATGGTAGTATATAAT	TAATTGTTGTTGCCGCCTCTG	pES_46
lncRNA-19	TATTTAAATTAGATTACTAACCCTTTGTTGTAGCTGGTG	AGGCTGCCGTTAGGGTGAGC	TGTGTTATCAACAAACAATTTATGAAATGGG	CAAATTTAAATAAGTAATTAAATAAACACTGGGCA	pES_49
Plasmid used for insertion of SV40 stop signal					
Candidates	**[1]Left homology arm primer 1**	**[2]Left homology arm primer 2**	**[3]Right homology arm primer 1**	**[4]Right homology arm primer 2**	**Plasmid #**
lncRNA-9	CTTCTGCCTTCTTCTGCAACGAG	CAGTCTTTCTTCTATTTTGTTTTTTTGTTTTTTTTTTTTTTGG	AAAAAGTAACCATTAGAGAAGTGGTTAAAGTG	AGTTGATTGCATTCAGGCTGGG	pES107c

[1] 'GGTGGCGGCCGCTCTAGAACTAGTG' was added to the 5' end of the left homology arm primer 1.
[2] 'ATTGAATTAGTCTCTAATTGAATTAGATCC' was added to 5' end of the left homology arm primer 2.
[3] 'TGAAACAATTAGGCAAAACCATGCGAAGCT' was added to 5' end of the right homology arm primer 1.
[4] 'ACGGTATCGATAAGCTTGATATCGAATTCC' was added to 5' end of the right homology arm primer 2.

Table 5: Typical PCR reaction condition

Temperature (°C)	Time (min)	Number of Cycles
94	2:00 min	1
94	0:30 min	20-40 cycles
55 to 60	0:20	
70	0:20-2:00 (20 sec per kb)	
70	1:00	1
4	∞	

Table 6: Oligonucleotides generated for this study

Primer Name	Sequence
3xP3-RFP_fwd	GGATCTAATTCAATTAGAGACTAATTCAATTAGAGCT
3xP3-RFP_rev	AGCTTCGCATGGTTTTGCCTAAT
pCDF4_fwd	GACACAGCGCGTACGTCCTTCG
pCFD4_rev	TCAGATGTATGTACGTCAACGGAAAACCATTGTC
pCDF3_seq_fwd	GAAAAGGTTAGCTCGCCAAG
pCDF3_seq_rev	GCATACGCATTAAGCGAACA
SV40terminator_fwd	GGATCTTTGTGAAGGAACCTTACTTCTG
SV40terminator_rev	GGATCCAGACATGATAAGATACATTGATGA
Donor construct sequencing primers	
pBSII_Donor_seq_1	CTGCAAGGCGATTAAGTTG
pBSII_Donor_seq_2	GGTTCGAAATCGATAAGCT
pBSII_Donor_seq_3	AGGATCCAAGCTTATCGAT
pBSII_Donor_seq_4	CCTTCAGGAAATGACAGCT
pBSII_Donor_seq_5	CCAGCGATCATTCATTTCG
pBSII_Donor_seq_6	GCACCCCAGGCTTTACACTTTA
Single fly sequencing primers (3xP3-RFP cassette insertion)	
3xP3_LHA_seq_rev	CCGCCCGATTGTTTAGCTTGTT
3xP3_RHA_seq_fwd	TCGTGATGGCGTGTTGAAAGGA
lncRNA-3	ACTCTTCTACGCATAAAATCAATATTCCCA
	GCGCTTTACCAACTTCACTTCAGCC
lncRNA-4	ACACTTACCAAATCCGAAATCATGA
	TCAGCTCGGTTCCATTGATAAGATG
lncRNA-5	ATGTCCCTTTTCGAGCGA
	TGGTAGCCCTCAGTACCG
lncRNA-6	CCAAACGCTTCTGTTCGAGCTT
	TGGAACTTGTAACTGTGGAATTGT
lncRNA-9	AGCAGCCCCATCTCCTTCACTG
	GCCAACTGACAGCCTACTTTCG
lncRNA-10	AGCAAAGCGTTCACACACAGAG
	AAGGCGATCATCATGTTTCGC
lncRNA-11	TGTTTGTCACGTTCCGCA
	TCTCTGGAAGTCGCCGAA
lncRNA-12	AGTTCCACCCCACGTCTAA
	TTGGCTCGAACTAACGGC
lncRNA-13	TCAGCTGAATGAGAGGGCT
	ATTGTGCGTGCTTTCGAC
lncRNA-14	CCACTGCGATCCGAAACA
	CAACGGGTGCTAAGGAGG
lncRNA-15	TGATTGTGACCGTTGCAGT
	ACTCATCATCGGCCAGGT
lncRNA-16	TCATAATGCTCAGCGGACGGAT
	TGCTCTTTAATTGAGCACCCTG
lncRNA-19	CGTTAGCCAATTTACGAGCGACT
	AGTTCCGCGATGCGATTTAAT
Single fly sequencing primers (SV40 insertion)	
lncRNA-9 (pES107c)	TGCCTAGCTGCGGATTAACATGCA
	AATTCAGGGGTCAAGGAACACC

Table 7: Oligonucleotides generated for qPCR

Primer Name		Sequence	
Corazonin (Crz)	CAACGGCCATCTCCATCGGG GTCTGGGATGGGCGCTGTTT ATCCCGAAAACAGCGCCCAT AACACATTGGGTTCGGCGGA GCCCCTCTTCCTCTTCACGC GCCTACGATCGCTGCTCCAA	Insulin-like peptide 2 (Ilp2)	TTGGCCCAAGGAACGCTCTG AACTGCAGGGGATTGAGGGC ATTCCACACAAGCGCGCCAT TCCGGACTTCAGCCAGGGAA CCCAAGGAACGCTCTGCAGT GCAGGGGATTGAGGGCGTC
Leucokinin (LK)	GCAGCTCTCCAAGTACCGCA CGCTGAGATGTGGGTGGTGG CGGTGGCCAGGATTCGGAC AGCTGGGATTTGATGCGGCA ACCGCTGGAGAAGATGGCAA GAGTCCGAATCCTGGCCACC	Insulin-like peptide 3 (Ilp3)	CGAAGACCGTTCCCTGCTGG TCGCAGTTGTCGTCATTGGGT ACCCAATGACGACAACTGCGA TGGCCCCGCACAATCATACA CGATTGGGGAAAGCACTCACG CGATGTATTCTGGCCCCGCA
Capability (Capa)	TCGCCAAGCCAGTGTTTCGT GCCTCTAATCCGTCGCGCAG TCTGCGCGACGGATTAGAGG CACTTCCTCAGCTCGGCGTC AGTGGTCTGGTGGCCTTCCC GTCCACACTGCGCTTTCCCA	Akh	GATTGGGGCAAGCGTTCGGT TACTCGCGGTGCTTGCAGTC GCCGTGCTCTTCATGCTGCT GCGGAAGATCTCGAGCAGCA GCGAAGTCCTCATTGCAGCC GCGGTGCTTGCAGTCCAGAA
Eclosion hormone (Eh)	TGCATTGCCCGCCATAAGTCA GAGGCAGGAAAGGGCACAGG TGTGCCTTAACAACTGCGTCCA CGAGCGCATTCAGAAACGGG AGCCCTTGATTTTGTGCACCT GCAGGAAAGGGCACAGGTCT	Tachykinin (Tk)	ACTGAAGAAGGCACCGCTGG TACCTCGCATGCCCGTGAAC GTTCACGGGCATGCGAGGTA ACGCTTGTAGTGCTGGTGGG CGGAGGGTAACTGGCTGGGA CAGGCTCTGCAGGACATCGG
short neuropeptide F precursor (sNPF)	GGTGTCCCTCCAGATGCAGC CGCCGAACCAACGCTTTTCC GGAAAAGCGTTGGTTCGGCG CGTAGTCGGCACCCAGATCG TCACCCTCAACAGCCAGCAG CCGCGTTGTTTGCCTCGAAC	AstA	TCCCCGGTAATTGGCCAGGA AGACCTCTTGCCCAGACCGA GTCTGGGACGACGGGCCTAT TCGGCTGCCAAGTAGTTCGC GAACTCCCTTCACGCCCACC TATAGGCCCGTCGTCCCAGA
FMRFamide (FMRFa)	TCCCGAGTCCAAGCCAGTCA CCACACTCCCATCCTGCGTG TTGTGCGCTCCGGGAAGATG CGCCGTTCTTCACCTGCTGA GTTCCTGCTGGCCCTGTACC AGAGATGGGGTGCCGGAACT	Hugin (Hug)	GCAGCATTGACTCCTGGCGA ACCGCACACTTGAGCACGTT GGAAATCACATCAGCGCCGG ATCGAATTCCTGGCCCTGCG CGCAGGGCCAGGAATTCGAT CGTCCAGTAGTCGCCAGGAG
Nplp1	TTATCGCCGGGCAGAGAAGC GCCCCGCCAAAACCACTACT GCAACTATGGCCCGTCTGCA CGGCGAGGACTTCAGAGCTC TCGCACGCTACAACTCGCAG GGAATCGCTTCTCTGCCCGG	Crustacean cardioactive peptide (CCAP)	TCAATGCCGAGCCAGCTGTG TTCGCGCTCCTCCAATTGCT TGCTTGCACTCCTGGCTTGT CGCTTTCGTCCACAGCCTGT CCTCGTATCCGCCATTCTCGC TGGCCTGCGACATGATTTGC
Nplp2	ATGGCCAAGCTCGCAATTTG CTCAACCTTCTTCGCGTCCA GTGTTCGCCCTGTTCGCCTT CGTTGAAATCACCCTGGGCCT CCTGTTCGCCTTGGCTCTGT GTCCAACGCCTTCAACTTTTCAA	Orcokinin	CGACCTTTCCACATCGCCGT ATGACCGCCACCAATGCTGT AACTTCATTCACGCAGCGCC CTTGCAAGGACTCGGGGCTC CGGCTGCCAATGGACAAACG GCGGAGGCCTTGTCGATCTC GTGCAGCCAGTCGGAGTTCA AGACGAGCTGGTTGAGGATGC

Table 8: *Drosophila* species embryonic development RAMPAGE time series (.sra)

Time after egg laying (hr)	*Drosophila ananassae*	*Drosophila erecta*	*Drosophila melanogaster*	*Drosophila pseudoobscura*	*Drosophila simulans*
0-1	GSM2365603	GSM2365709	GSM2365625	GSM2365648	GSM2365687
1-2	GSM2365604	GSM2365710	GSM2365626	GSM2365649	GSM2365688
2-3	GSM2365605	GSM2365711	GSM2365627	GSM2365650	GSM2365689
3-4	GSM2365606	GSM2365712	GSM2365628	GSM2365651	GSM2365690
4-5	GSM2365607	GSM2365713	GSM2365629	GSM2365652	GSM2365691
5-6	GSM2365608	GSM2365714	GSM2365630	GSM2365653	GSM2365692
6-7	GSM2365609	GSM2365715	GSM2365631	GSM2365654	GSM2365693
7-8	GSM2365610	GSM2365716	GSM2365632	GSM2365655	GSM2365694
8-9	GSM2365611	GSM2365717	GSM2365633	GSM2365656	GSM2365695
9-10	GSM2365612	GSM2365718	GSM2365634	GSM2365657	GSM2365696
10-11	GSM2365613	GSM2365719	GSM2365635	GSM2365658	GSM2365697
11-12	GSM2365614	GSM2365720	GSM2365636	GSM2365659	GSM2365698
12-13	GSM2365615	GSM2365721	GSM2365637	GSM2365660	GSM2365699
13-14	GSM2365616	GSM2365722	GSM2365638	GSM2365661	GSM2365700
14-15	GSM2365617	GSM2365723	GSM2365639	GSM2365662	GSM2365701
15-16	GSM2365618	GSM2365724	GSM2365640	GSM2365663	GSM2365702
16-17	GSM2365619	GSM2365725	GSM2365641	GSM2365664	GSM2365703
17-18	GSM2365620	GSM2365726	GSM2365642	GSM2365665	GSM2365704
18-19	GSM2365621	GSM2365727	GSM2365643	GSM2365666	GSM2365705
19-20	GSM2365622	GSM2365728	GSM2365644	GSM2365667	GSM2365706
20-21	GSM2365623	GSM2365729	GSM2365645	GSM2365668	GSM2365707
21-22	GSM2365624	GSM2365730	GSM2365646	GSM2365669	GSM2365708
22-23	-	GSM2365731	GSM2365647	GSM2365670	-
23-24	-	-	-	GSM2365671	-
24-25	-	-	-	GSM2365672	-
25-26	-	-	-	GSM2365673	-
26-27	-	-	-	GSM2365674	-
27-28	-	-	-	GSM2365675	-
28-29	-	-	-	GSM2365676	-

Table 9: GEO Accession numbers linked to modENCODE RNA-seq

Time after egg laying (hr)	SRA accession files
0-2	SRR767606, SRR1197337, SRR1197370
2-4	SRR767626, SRR1197336, SRR1197368
4-6	SRR767609, SRR1197338
6-8	SRR767610, SRR1197333
8-10	SRR767615, SRR1197335
10-12	SRR767616, SRR1197334, SRR1197367
12-14	SRR767613, SRR1197332, SRR1197369
14-16	SRR767618, SRR1197331
16-18	SRR767605, SRR1197330, SRR1197365
18-20	SRR767622, SRR1197327, SRR1197363
20-22	SRR767620, SRR1197329, SRR1197364
22-24	SRR767625, SRR1197328, SRR1197366

Table 10: GEO Accession numbers for RAMPAGE DATA

Time after egg laying (hr)	SRA accession
0-1	SRR424683
1-2	SRR424684
2-3	SRR424685
3-4	SRR424686
4-5	SRR424687
5-6	SRR424688
6-7	SRR424689
7-8	SRR424690
8-9	SRR424691
9-10	SRR424692
10-11	SRR424693
11-12	SRR424694
12-13	SRR424695
13-14	SRR424696
14-15	SRR424697
15-16	SRR424698
16-17	SRR424699
17-18	SRR424700
18-19	SRR424701
19-20	SRR424702
20-21	SRR424703
21-22	SRR424704
22-23	SRR424705
23-24	SRR424706

Chapter 3 |

Pilot screen for lncRNAs that function during *Drosophila* embryogenesis

Since this project was performed in collaboration with Drs. Philippe Batut & Thomas Gingeras (Cold Spring Harbor Laboratory, USA), I will describe our individual contributions. The bioinformatic analysis of RAMPAGE data was performed by Dr. Philippe Batut from Dr. Thomas Gingeras's lab without contributions from me. The microinjection service and subsequent isolation and backcrossing of flies to a balancer was performed by the fly facility at the Department of Genetics, University of Cambridge without contributions from me. For the purpose of this thesis, I have designed and validated the plasmids and performed all the other experiments. I have also analysed the data to generate the figures in this chapter.

3.1 Introduction

Following the advent of high throughput sequencing technologies, it is now well established that eukaryotic gene expression is highly pervasive in all organisms and is not limited to the protein-coding regions of the genome (Okazaki *et al.,* 2002; Carninci *et al.,* 2005; Birney *et al.,* 2007; Derrien *et al.,* 2012). The transcriptome harbours only ~2% of protein-coding genes while for the vast majority of the transcriptome is made into ncRNAs. Among these ncRNAs are non-coding transcripts known as lncRNAs and characterised by a length above 200 nucleotides and the lack of apparent protein-coding potential (Guttman *et al.,* 2009; Derrien *et al.,* 2012). From large-scale transcriptome analyses, these long transcripts have been identified to be a major component of the ncRNA cohort, yet relatively little is known about their functions during development. Despite the emerging evidence of distinct lncRNAs with diverse physiological functions, very few of these have been studied experimentally in any detail and the functions of only a handful of lncRNAs have been described to date (Moran *et al.,* 2012).

Previous characterisation of lncRNA functions was performed primarily in cell lines and thus only limited direct genetic evidence of their functional *in vivo* implications is available. Earlier studies on various animal models have demonstrated compelling evidence for the requirement of lncRNA transcripts for organ development, organismal survival and normal physiological functions (Grote *et al.,* 2013; Sauvageau *et al.,* 2013; Hardiman et al., 2002; Li *et al.,* 2012). However, there are not many examples in which consequences of lncRNA loss of function were systematically assessed and thus making it challenging to perform subsequent analysis to understand the impact of the lncRNA deletion (Sauvageau et al., 2013). Most of these screens have been carried out on cells rather than examining them in vivo.

In recent years, lncRNAs have been identified and catalogued at different stages of development from embryo to a fully-grown adult in different model organisms. Focusing on the events that take place during embryonic development in mouse, lncRNAs were found to have a role in specifying the different lineages derived from the three germ layers including from neural, mesodermal and endodermal origin (Poirier *et al.,* 1991; Jiang *et al.,* 2015; Winzi *et al.,* 2018). They were suggested to be important regulators during embryo morphogenesis and could potentially influence the development of embryos and organs when they are mutated. Examples of lncRNAs that were first discovered to have a role during embryogenesis include lncRNA *H19* and *Xist* (Leighton *et al.,* 1995; Rancourt *et al.,* 2013; Penny *et al.,* 1996; Marahrens *et al.,* 1997). Subsequent efforts have identified lncRNAs such as *Fendrr, Braveheart, Kcnq1ot1* that were involved in organogenesis during mouse embryonic development (Grote *et al.,* 2013; Klattenhoff *et al.,* 2013; Korostowski *et al.,* 2012). For example, the lncRNA *Fendrr* localises to the nucleus and plays an essential role in the differentiation of cardiovascular tissues (Schmitz *et al.,* 2016). In the absence of *Fendrr,* the heart and the body wall of the embryo were not generated due to the inability of the cells from lateral mesoderm to undergo cellular differentiation which eventually resulted in embryonic lethality (Grote *et al.,* 2013; Grote and Herrmann; 2013).

The vast amount of developmental genetics and functional studies carried out throughout the years have made the *Drosophila* embryo one of the most understood examples of transcriptional regulation in developmental biology. The fruit fly has been extensively studied for over a century and it has many characteristics which makes it an ideal organism for the elucidation of lncRNA function during embryogenesis. The functional annotation of lncRNA transcripts in the *Drosophila* genome is an essential first step to appreciate the molecular and

cellular mechanisms that underpin the diverse roles that lncRNAs play in regulating various aspects of development. One of the goals of the MODENCODE Consortium was to annotate the *Drosophila* genome and identify functional transcriptional start sites (TSS) (Roy *et al.,* 2010). The expression pattern of lncRNAs has previously been reported to exhibit temporal dynamics during development or cellular differentiation and the choice of a lncRNA candidate for further in-depth study is often associated with a specific defect (Sauvageau *et al.,* 2013; Yan *et al.,* 2015). Of the few lncRNA mutant flies generated to date, most involved previously studied and classic examples such as *rox1*, *rox2*, *bxd* (Petruk *et al.,* 2006; Franke and Baker, 1999; Meller *et al.,* 1997), leaving the more recent large-scale RNA-Seq-derived catalogues of lncRNAs largely unexplored. In this project, I decided to explore the biological role of novel lncRNA candidates expressed during specific embryonic developmental stages in *Drosophila* as little is known regarding lncRNA function during this period.

To this aim, we decide to use previously existing datasets generated by RNA Annotation and Mapping of Promoters for Analysis of Gene Expression (RAMPAGE) to select candidate lncRNAs for a functional genetic screen (Batut *et al.,* 2017). These libraries had the advantage of a high temporal resolution (1-hour resolution throughout embryonic development) and deep sequence coverage across 5 *Drosophila* species, which allowed us to select for lncRNA candidates that were expressed during *Drosophila* embryogenesis with high confidence (Batut *et al.,* 2017). Moreover, sequence conservation across species is usually interpreted as a sign of functional importance. As lncRNAs are poorly conserved at the sequence level, the RAMPAGE datasets were used to identify transcriptional start site clusters (TSCs) not associated with known coding genes or small RNAs and that also showed syntenic conservation across all 5 species and were transcriptionally active in all of them, suggesting critical functional roles for these lncRNAs. A stringent set of 631 lncRNA that are conserved across all 5 *Drosophila* species was identified (Batut *et al.,* 2017).

As it is time-consuming and costly to perform *in vivo* studies, we needed to further reduce this list of lncRNAs to a manageable number to functionally assess. Dr. Philippe Batut applied additional selection criteria to further narrow down this list and eventually he manually curated a small subset of twenty-two lncRNAs candidates, that I have then brought forward for the genetic screen. Besides 20 novel lncRNAs our list also included two well-known *Drosophila* lncRNAs, *bxd* and *rox1*, which fulfilled all our criteria and served as positive controls for our functional studies (Batut *et al.,* 2017; Petruk *et al.,* 2006; Meller *et al.,* 1997). CRISPR/Cas9-mediated deletion was used to specifically target the lncRNA

promoter regions via a pair of guide RNAs (gRNAs) (Port *et al.,* 2014). Using 5′- and 3′-flanking gRNAs that surround the transcriptional start site of each lncRNA candidate, large stretches of DNA can be deleted, thus allowing efficient disruption of lncRNA loci. The promoter deletion strategy was chosen as it will inhibit lncRNA transcription and therefore result in a phenotype that is caused by the lack of lncRNA (Gratz *et al.,* 2014; Bottcher *et al.,* 2014; Port *et al.,* 2014). To date, I have made use of the CRISPR/Cas9 technology to successfully generate 13 lncRNAs mutants out of the 22 candidates that were used for further analysis. CRISPR was also attempted for other 9 lncRNAs but no founders were successfully generated after 2 attempts.

Here, I am reporting the initial characterisation of these new knockout strains which included testing the mutants for some of the phenotypes that have been commonly studied in lncRNA loss-of-function studies such as viability, fertility, locomotion and developmental process (patterning of the body plan) (Sauvageau *et al.,* 2013; Goudarzi *et al.,* 2019; Li *et al.,* 2012; Pease *et al.,* 2013). In this chapter, I will describe the observed phenotypes for the 13 lncRNA mutants that I have generated. While the majority of the lncRNA mutants do not have any observable developmental abnormalities, abrogation of two of the lncRNAs resulted in a homozygous lethal phenotype. As a whole, these data provided us with new evidence of the roles that lncRNAs play in *Drosophila* development and physiology.

3.2 Results

3.2.1 Selection of lncRNA candidates for pilot screen

The complete development of a multicellular organism is coordinated by a precise spatio-temporal expression of both protein-coding and non-coding genes. A robust regulation of gene expression from embryonic development to a complete adult requires an accurate and coordinated expression of both classes of molecules. The discovery and characterisation of transcriptional regulatory elements such as promoters and enhancers improved our understanding of eukaryotic transcription and regulation of gene expression in a complex genome. In recent years, the identification and characterisation of transcriptional start sites (TSSs) and promoters at the genome-wide level has been made possible with the development of high throughput sequencing technologies in combination with new advancement in TSS profiling methods. The use of RNA-seq methodology alone has its own caveats as it is

inaccurate at reconstructing the 5′ ends of transcripts; moreover, proper detection of lowly-abundant transcripts such as lncRNAs requires an in-depth sequencing to be performed. A technique known as RAMPAGE (Batut *et al*., 2013) was developed to overcome these constraints. In this technique, the use of a template switching reverse transcriptase coupled to cap trapping by biotinylation and pulldown of capped RNA molecules allows to select 5' complete cDNA replicas of capped RNAs for downstream library preparation. This system therefore enables a user to sequence 5'-complete cDNAs and allows the precise identification of TSSs and at the same time reveals the promoter activity of each gene at genome-wide level (Batut *et al*., 2013; Batut *et al*., 2017). RAMPAGE was first employed to characterise the promoter architecture throughout the 36 stages of the life cycle of *Drosophila melanogaster*. An extensive usage of transposon-driven alternative promoters was discovered during this analysis and it was shown to drive the expression of various developmental genes at different time points during development (Batut *et al*., 2013). A brief protocol of the RAMPAGE method can be found in Material and Methods and in Figure 12 and a more detailed methodology has been published (Batut *et al.,* 2013, Batut and Gingeras, 2013).

In order to examine the transcriptional regulation during fly embryonic development at the genome-wide level, RAMPAGE was used to analyse the activities of promoters across five *Drosophila* species (*Drosophila melanogaster, Drosophila simulans, Drosophila erecta, Drosophila ananassae, Drosophila pseudoobscura)* (Batut *et al*., 2017). In this manuscript, it was already established that close to 4,000 transcriptional start sites likely representing lncRNA promoters can be identified as functionally active during fly embryogenesis, with some of these promoters showing conserved activities across all five *Drosophila* species. We therefore took advantage of these already existing datasets and extended the previous analysis to select 20 novel lncRNAs of potential relevance during *Drosophila melanogaster* embryonic development for further functional studies. A flowchart of the experimental design can be found in Figure 13. The data that were used to select candidate lncRNAs were generated and analysed entirely by Dr. Philippe Batut (as part of a collaboration with the laboratory of Dr. Thomas Gingeras at Cold Spring Harbor Laboratory, USA).

The RAMPAGE transcriptomic datasets were ideal for this purpose because they offer both a high temporal resolution (1 hr) as well as a high sequence coverage (137-180 million reads per species) for the period of embryonic development (Figure 13). Moreover, the comparison across different species allowed us to focus on lncRNAs that are syntenically conserved and expressed in all 5 species with a similar expression pattern, suggesting

functional importance. A total of 120 RAMPAGE libraries were previously generated for the five species and the different timepoints, each library was aligned against its reference genome using STAR aligner (Dobin *et al.,* 2013) and the expression of annotated genes was quantified using Cufflinks. A novel peak-calling algorithm for TSS discovery was first used generate individual promoters by clustering neighbouring TSS into TSS clusters (TSC) followed by an evaluation of the activity of each these promoters (Batut *et al.,* 2013, Batut and Gingeras, 2013). A total of $2.2x10^4$ to $2.7x10^4$ high-confidence TSCs were identified for each species during the analysis. An excellent reproducibility (Pearson $R^2 = 0.95$) for TSCs with maximum expression of ≥ 25 reads per million (RPM) and reproducibility maximum expression ≥10 RPM (Pearson $R^2 = 0.92$) was achieved when the activity of the promoters were compared across the biological replicates for the *Drosophila melanogaster* time series. Quantification across other *Drosophila* species also yielded similar results. In addition, the use of paired-end sequencing has allowed direct association of the 5'end of a transcript with its associated mRNA (Batut *et al.,* 2017).

Drosophila embryonic development time course is widely different across the various species. To be able to compare expression profiles across species, post-synchronisation of expression profiles was measured as follows. First, the raw read counts per TSC were normalized to library size for each sample, and subsequently expressed in reads per million uniquely mapped reads (i.e., RPM). Those RPM counts were then normalized to a Z-score across each developmental time series for each individual TSC. Following these normalization steps, the similarity of these Z-score "profiles" across 2 species was defined for each TSC as the Pearson R^2 for the 2 expression vectors being compared. Finally, the overall profile conservation metric used is the average Pearson R^2 across all possible species pairs. To compare the overall similarity between genome-wide expression profiles for the different species, the Z scores for all-time series for each species were registered to one another. Overall tightly conserved expression profiles were observed for genes with one-one orthologs across all species studied, even if the amount of conservation could be quite variable among different genes (Batut *et al.,* 2017).

During this analysis, thousands of novel promoters that could potentially drive the expression of long non-coding transcripts were found. To explore the functional relevance and pervasiveness of lncRNAs expression, the application of a phylogenetic framework allows the conservation of each lncRNA to be examined. In addition, RAMPAGE can pinpoint the TSSs with single-base resolution allowing for the identification of conserved

sequence patterns. Starting with a list of 25,426 TSCs that have been confidently detected in *Drosophila melanogaster* through RAMPAGE, it was established that 3,682 of these TSCs could not be linked functionally to any Flybase annotation (protein-coding or small RNA gene). An additional 291 TSCs were identified to be associated with Flybase-annotated lncRNA genes, which brings the total number of putative lncRNA TSCs to 3,973. Despite the fact that lncRNA sequences are poorly conserved across species, both the primary sequence and the expression pattern of the identified lncRNA promoters was found to be conserved as many of them were found to be under a high degree of selective pressure compared to protein-coding genes. It is hypothesised that these lncRNAs could potentially be playing a broad diversity of functions due to the similar expression patterns when compared to those of protein-coding genes. Furthermore, for the selection of a rigorous set of lncRNAs, a stringent boundary on the true number of conserved lncRNA promoters were placed. Emphasis was placed on those TSCs that were shared between all five *Drosophila* species and many TSCs have been filtered out solely because of the poorer quality of the genome assemblies. The 631 TSCs that were confidently aligned to all genomes, as well as transcriptionally active in all 5 species made up the list of potential lncRNA candidates (Figure 13). A high degree of promoter conservation was found in these TSCs, suggesting strong lineage-specific selective constraints.

For the purpose of this study, initial emphasis was placed on the most highly transcribed TSCs in *Drosophila melanogaster*. Starting from the top of the list, the TSCs were ranked according to the conservation of their expression profiles across species. The final step of lncRNA candidate selection was based on a few criteria and was largely manual. We have selected TSCs with clean RAMPAGE signal, high-quality peak calls (minimum expression of at least 10 reads per million [RPM]), transcripts that were clearly independent of neighbouring units and preferentially those that were active during a short developmental window. The candidates were ranked accordingly and the first 20 candidates were selected. Two well-studied lncRNAs in *Drosophila* that satisfied all of these criteria were also included to serve as positive control. These are lncRNA-5 (*FBgn0019661_05, rox1*) and lncRNA-6 (*FBgn0020556_01*, *bxd*). The biological function of the 20 potential lncRNA candidates will be studied in this project (Table 11).

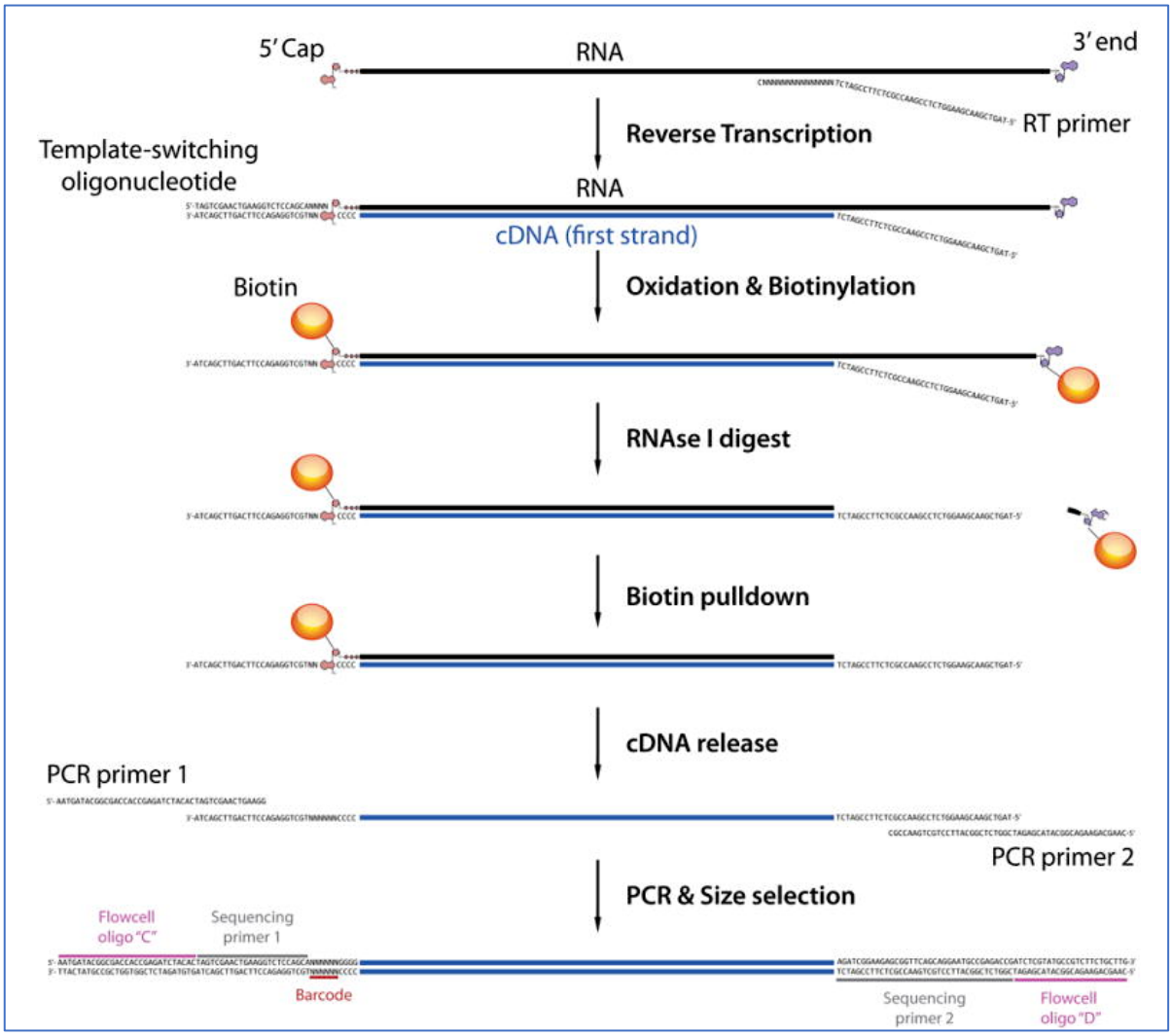

5′ Cap
RNA
3′ end
RT primer
Reverse Transcription
Template-switching oligonucleotide
RNA
cDNA (first strand)
Biotin
Oxidation & Biotinylation
RNAse I digest
Biotin pulldown
cDNA release
PCR primer 1
PCR primer 2
PCR & Size selection
Flowcell oligo "C"
Sequencing primer 1
Barcode
Sequencing primer 2
Flowcell oligo "D"

Figure 12: A brief summary of the RAMPAGE library preparation protocol
High quality, ribosome-depleted total RNA extracted from embryos is reverse-transcribed (RT) to synthesize first strand cDNA using a custom RT primer having an Illumina adaptor sequence overhang. During the reverse transcription reaction, amplifiable 5′-complete cDNAs are generated as a result of template switching by the reverse transcriptase induced by the presence of a template-switching oligo (TSO) containing the other illumina primer, a unique barcode and three riboguanosine at its 3' end. Oxidation and biotinylation of the 5'-m7G caps were used as part of the cap-trapping approach to bind the hybrid RNA/cDNA molecules to streptavidin beads. PCR amplification and size selection is carried out on the single-stranded cDNA released from the streptavidin-coated beads. The final RAMPAGE library is generated using custom oligonucleotides and can be directly sequenced on Illumina platforms. [Image taken from Batut *et al.,* 2013]

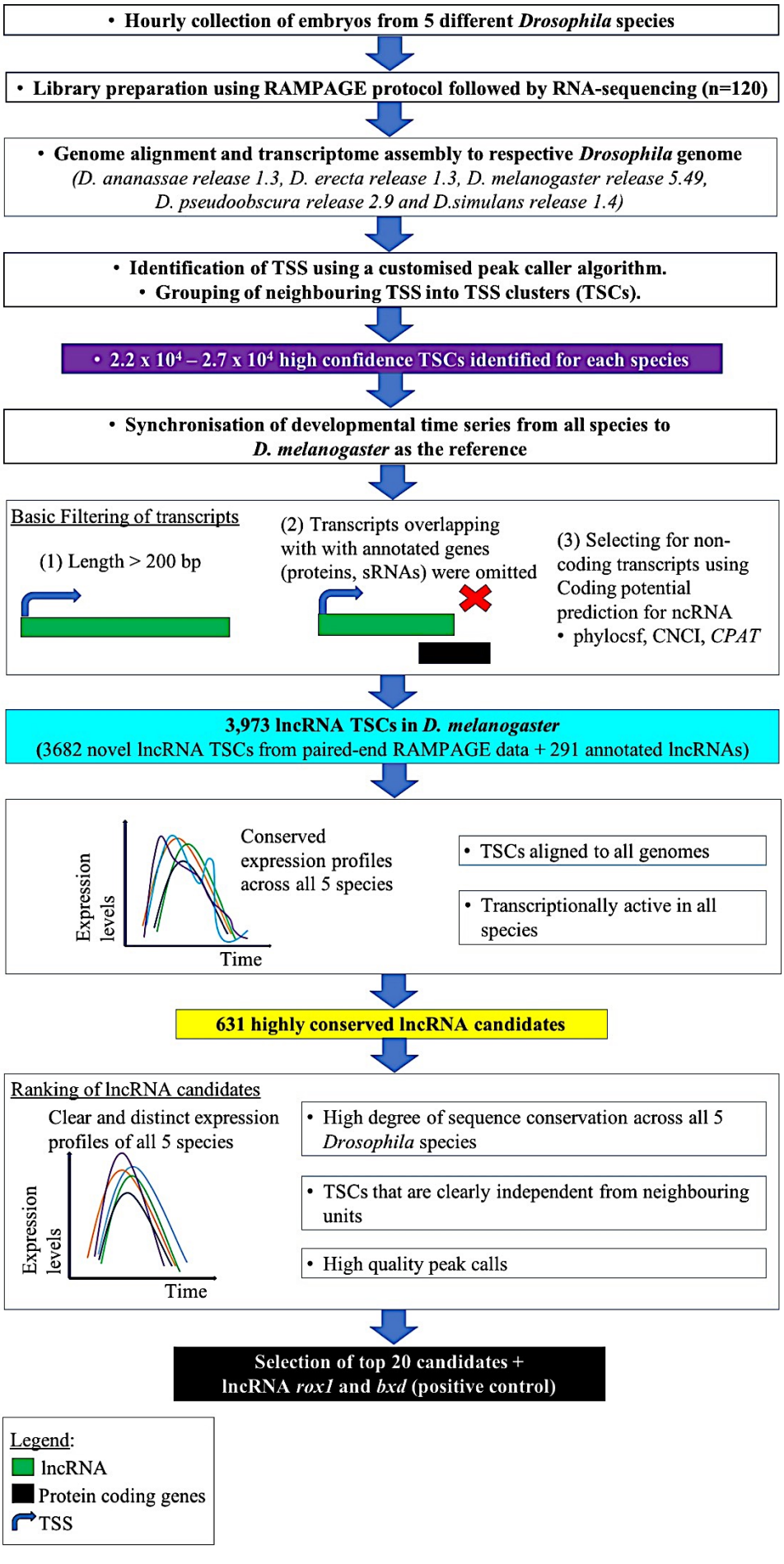

• Hourly collection of embryos from 5 different Drosophila species
• Library preparation using RAMPAGE protocol followed by RNA-sequencing (n=120)
• Genome alignment and transcriptome assembly to respective Drosophila genome
(D. ananassae release 1.3, D. erecta release 1.3, D. melanogaster release 5.49, D. pseudoobscura release 2.9 and D.simulans release 1.4)
• Identification of TSS using a customised peak caller algorithm.
• Grouping of neighbouring TSS into TSS clusters (TSCs).
• 2.2 x 10⁴ – 2.7 x 10⁴ high confidence TSCs identified for each species
• Synchronisation of developmental time series from all species to D. melanogaster as the reference
Basic Filtering of transcripts
(1) Length > 200 bp
(2) Transcripts overlapping with with annotated genes (proteins, sRNAs) were omitted
(3) Selecting for non-coding transcripts using Coding potential prediction for ncRNA
• phylocsf, CNCI, CPAT
3,973 lncRNA TSCs in D. melanogaster
(3682 novel lncRNA TSCs from paired-end RAMPAGE data + 291 annotated lncRNAs)
Expression levels
Time
Conserved expression profiles across all 5 species
• TSCs aligned to all genomes
• Transcriptionally active in all species
631 highly conserved lncRNA candidates
Ranking of lncRNA candidates
Clear and distinct expression profiles of all 5 species
Expression levels
Time
• High degree of sequence conservation across all 5 Drosophila species
• TSCs that are clearly independent from neighbouring units
• High quality peak calls
Selection of top 20 candidates + lncRNA rox1 and bxd (positive control)
Legend:
lncRNA
Protein coding genes
TSS

Figure 13: Pipeline for the identification of potential lncRNA candidates.
A genome-wide analysis of the core promoter activities during *Drosophila* embryogenesis was performed using RAMPAGE. Hourly collection of embryos from five different *Drosophila* species allowed the examination of different promoter activity at various stages of development. The RAMPAGE reads were aligned to the respective *Drosophila* genomes with the 5' end from each properly-paired read predicted as a potential transcriptional start site (TSS). Neighbouring TSS were grouped to form TSS clusters (TSC) and the raw read counts per TSC were normalized to library size for each sample, and subsequently expressed in reads per million uniquely mapped reads (i.e., RPM). A Z-score was obtained from the normalisation of RPM counts for each developmental time series for each individual TSC. A total of 2.2×10^4 to 2.7×10^4 TSCs were confidently identified from all five *Drosophila* species. Using all TSCs confidently detected in *Drosophila melanogaster* Replicate 1 (25,426 TSCs), a total of 3,973 potential lncRNA TSCs not linked to any Flybase annotation were identified using paired-end RAMPAGE data. This number includes the 291 TSC associated with Flybase-annotated lncRNA genes. A stringent selection pipeline was chosen to identify TSCs that were confidently aligned to all genomes and also transcriptionally active in all 5 species. Following this analysis, 631 conserved lncRNA TSCs were confidently identified. Focusing on the most highly transcribed TSCs in *Drosophila melanogaster,* TSCs were ranked according to the conservation of their expression profiles across species. The final step of candidate selection was largely manual as we have picked TSCs with clean RAMPAGE signal, high-quality peak calls, transcripts that were clearly independent of neighbouring units and preferentially active during a short developmental window. The top 20 candidates were selected for further genetic screen to understand their functional relevance. 2 known lncRNAs that satisfy all the criteria above (*roX1 and bxd*) were added to the list as positive controls (Batut *et al.,* 2017).

3.2.2 Analysis of lncRNA candidates in the new *Drosophila* genome assembly (dm6)

With the introduction of the new *Drosophila* genome assembly (dm6, BDGP Release 6 plus ISO1 MT), the "lift over" function on the UCSC genome browser was used to locate the new coordinates for each of the lncRNA candidates (Table 12). The previous assembly of the *Drosophila* melanogaster genome (dm3, BDGP Release 5) was first used as the latest reference genome assembly (dm6, BDGP Release 6 plus ISO1 MT) was only released after the RAMPAGE analysis. During the changeover, the NCBI RefSeq and Ensembl databases were also updated and the number of annotated lncRNAs identified went up from 9 to 17. The functions of the newly annotated lncRNAs remained unknown. Noteworthily, lncRNA-8, which was previously identified as a non-coding RNA, had been annotated as a protein-coding gene and will be removed from this study. Therefore, the total number of lncRNAs for this project became 21 instead of 22. The Coding-Non-Coding Index (CNCI) was applied to evaluate the coding potential of the 21 potential lncRNA candidates and all of them had a CNCI score of < 0, indicating that they were likely non-coding transcripts (Table 12). LncRNA-4, lncRNA-19 and lncRNA-22 had the best CNCI score (-0.0004096) of all the candidates.

The TSS data generated using RAMPAGE was compared to the CAGE reads tracks provided by the EPD team at the Swiss Institute of Bioinformatics (Dreos *et al.,* 2013). These tracks displayed information of putative TSS based on the 5′ end of transcripts (Chen *et al.,* 2014). In all cases, I did not observe any major inconsistencies between the TSS identified by CAGE and RAMPAGE. Although the CAGE reads do not align perfectly with the data from RAMPAGE all the time, transcripts for the lncRNA candidates were captured by both methods. The CAGE tracks provided by EPD for the dm6 genome assembly also contained additional information about the adult tissue (head, digestive system, testes, ovaries) and stage-specific (embryo, larvae, pupae) information for each transcript. This information provide us with the potential role that the lncRNA is involve in. For example, if a lncRNA is expressed in testes or ovaries, it is probably not involved in courtship behaviour All of the candidate lncRNAs transcripts were captured by CAGE during embryogenesis. However, the majority of them (16 out of 21) are not restricted to embryonic development and the transcripts were also found to be expressed during larvae and/or pupae development. Some of the candidates had tissue-specific expression and this information could give us clues as to the function of the lncRNA (Table 13). Some of the candidates had broad expression patterns and the transcripts was captured in almost all of the tissues (lncRNA-1, lncRNA-5, lncRNA-6)

whereas some had a specific tissue expression. I found that a subset of the candidates in our list had a specific expression in the head indicating a possible neurological or behavioural function (lncRNA-3, lncRNA-4, lncRNA-9, lncRNA-10, lncRNA-11, lncRNA-14, lncRNA-15, lncRNA-21, lncRNA-22).

3.2.3 Generation of fly lines for somatic CRISPR

Generation of dual-gRNA expression constructs

In order to determine if the disruption of a lncRNA candidate would have detrimental effects during development, I took advantage of the CRISPR/Cas9 system as a genome-engineering tool to generate lncRNA mutants in *Drosophila.* A dual-gRNA expression construct strategy was utilized to target the lncRNA candidate of interest, as this would allow a defined, large genomic region to be deleted. It is relatively easy to create knockouts of protein-coding genes because a small deletion or insertion will disrupt the open reading frame, resulting in a non-functional gene product. Conversely, non-coding genes, especially lncRNAs, are very different. A small deletion or insertion may not necessarily result in a functional knockout as these lncRNAs lack open reading frames, some of these molecules have multiple functions which are likely connected to their three-dimensional structure and therefore it is extremely difficult to predict which parts of their primary sequence are essential.

I have tested different strategies to achieve the production of the lncRNA mutants. A pair of gRNA target sequences (20-nt) was selected for each lncRNA using online tools (http://www.flyrnai.org/crispr/) and cloned into the pCDF4 expression vector. The *D. melanogaster* genome assembly (dm3, BDGP Release 5) provided by the Berkeley *Drosophila* Genome Project (BDGP, http://www.fruitfly.org/) was used during the design of the gRNAs. The protocol and expression vector used (pCDF4) were similar to those previously described for *Drosophila* (Port *et al*., 2014). The pCDF4 vector contains two promoters – a *U6-1* promoter and a *U6-3* promoter – which drive the expression of each gRNA, respectively. Each genomic target sequence also harbours an NGG protospacer adjacent motif (PAM), which is required for Cas9 cleavage (Bassett *et al.,* 2013). Primers were designed using the vector as a template. Each forward and reverse primer contains a proto-spacer sequence, an overlapping region with the promoter and an overlapping region with the gRNA core sequence (Figure 14). Both PCR fragments and digested vector were run

on a gel to check for the correct product. The size of the PCR fragments should be ~600 bp while the size of the digested vector should be ~6.4 kb (Figure 14). PCR products for all lncRNAs yield products of the same size and are not shown. The final construct was verified via sanger sequencing. The genomic coordinates for each of the gRNA can be found in Chapter 2 (Table 3). The genomic region excised for each lncRNA is shown in Supplementary Figures 5-16.

Mutagenesis using somatic transgenesis produces mosaic patterns

The dual-gRNA plasmid targeting the individual lncRNAs or the white gene, were injected into *Drosophila* embryos (attP40 GFP- RFP-: y[1] w* M{vasint.Dm}ZH2A;; attP40) using the phage phi31 site-specific integration system to establish stable fly lines. The white gene is found on chromosome X and its mutation results in a visible eye-colour change that is easily detectable in adult flies. In addition, single copy mutations will produce a visible eye phenotype in male flies. The dual-gRNA stable lines can then be crossed to appropriate Cas9-expressing lines to generate mutants. I chose the strategy of separating the dual-gRNA and Cas9 expression in different lines because it guarantees better control of the system and potentially allows to study lethal mutants by using flies with conditional or tissue-restricted Cas9 expression. For the first attempt I have decided to assess somatic mutagenesis by utilising for the crosses flies expressing Cas9 in somatic tissues. As a pilot test for the feasibility of this approach, I used the white gene to assess the penetrance of the CRISPR/Cas9 system. Crosses between dual-gRNA-white flies and Act5c-Cas9 expressing flies resulted in the following phenotypic outcome. Analysis of the offspring revealed that ~25 % of the G1 had completely white eyes and ~25 % with wild-type eyes. The remainder (~50 %) produced varying degrees of mosaic eyes, with some eyes showing small red patches whereas others have large patches (Figure 14). The result of this preliminary experiment showed that only ¼ of the progeny generated by somatic mutation was fully mutant (white eyes) and all the rest of the flies were either a mosaic of mutant and wild-type cells or were completely wild-type. The somatic mutation strategy relies on nonhomologous end joining (NHEJ) repair. The overall low efficiency and a high probability of generating mosaic flies, makes it difficult to assess potential phenotypes. Moreover, the absence of an easy to follow marker was also a drawback of this initial strategy as significant time and effort would be needed to identify full lncRNA mutants in this approach. I therefore decided not to proceed further with this strategy and develop a more efficient approach to obtain full mutants and

easily select the mutant progeny. Nevertheless, the flies generated could potentially still be useful for conditional knockouts, if the mutant cells or tissues are identified.

3.2.4 Genetic deletion of large genomic fragment using Cas9-HDR

To overcome some of the limitations of the somatic transgenesis method described above, I decided to test the use of homology-directed repair (HDR) and germline-driven Cas9 as an improved strategy to generate lncRNA mutants as it allows for the germline mutations which gets rid of the mosaic issue. Having a mosaic pattern would be a big problem because it is unknown if the phenotype will be observable if only part of the cells is mutated. As the functions of the novel lncRNAs were unknown, I required a method able to reliably identify the geno- and phenotypes. Cas9-HDR was used to induce double-stranded breaks (DSBs) at the target loci and subsequent repair by HR using a donor construct containing a reporter for visual scoring. When the dual gRNAs guide the Cas9 protein to the target sites and cleave both DNA strands upstream of the PAM sequence, the DNA repair pathway is activated. If a donor plasmid containing homology arms close to the DSB site is present, it will act as repair template for HDR generating a knock-in of the repair template which might be designed to incorporate a reporter gene (Gratz *et al.,* 2014). In order to replace the segment between the pair of gRNAs through HDR, a cocktail containing a donor construct with the appropriate homology arms and a selective marker and its respective dual-gRNA expression vector was injected into embryos expressing Cas9 in germ cells. Each donor construct contained a 3xP3 promoter-driven red fluorescence protein (RFP) that is flanked by a pair of homology arms (approximately 1-1.5 kb in length each). In the event of a successful insertion, the RFP+ marker would replace the excised region by HDR and be visible in the adult eyes (Figure 16). Similar to the gRNAs, the homology arms were designed using the *D. melanogaster* genome assembly (dm3, BDGP Release 5) and visualised on the UCSC genome browser.

In addition to the use of HDR to insert a visible marker to easily identify mutant founders, we decided to use flies expressing Cas9 in the germline to ensure the recovery of stable lncRNA mutant lines. For lncRNAs that were on the X chromosome, the plasmid pool was injected into embryos expressing Cas9 in the germline (TH_attP2: y[1] sc[1] v[1];; {y[+t7.7] v[+t1.8]=nanos-Cas9}attp2). The rest of the pools were injected into Cas9 expressing embryos under the control of vasa regulatory sequences (BL55821: y[1] M{vas-Cas9.RFP-}ZH-2A w[1118]/FM7a, P{w[+mC] = Tb[1]}FM7-A). For each round of microinjection, the pool of gRNA and donor construct mix was injected into ~200

dechorionated embryos before cellularisation could take place (Figure 15). Staff at the Department of Genetics Fly facility performed the microinjection and subsequent isolation and backcrossing of RFP-positive flies to a balancer. The expression of the 3xP3-RFP cassette can be observed in larvae as well as in adult stages, thus flies with successful integration can be identified using a fluorescence dissecting scope.

In previous studies, the targeting efficiency of CRISPR/Cas9 have been reported to fluctuate significantly depending on the locus (Bassett *et al.,* 2013; Ren *et al.,* 2014; Gratz *et al.,* 2014; Port *et al.,* 2014). For the candidates that were successfully established, the success rate of genomic integration of the 3xP3-RFP cassette was low to moderate. Of the injected eggs that were viable for each lncRNA candidate, approximately 5 % of all embryos (200 embryos) that were co-injection with a dual-gRNA plasmid and donor construct had successful integration. On the other hand, only 0.5 % of all the progeny were RFP+. The majority of mutants that were generated required two rounds of transformation in order to have positive founders. The plasmids had to be re-purified, quantified and pooled for the second round of embryo injection. 13 lncRNA mutants were created in this study whereas none of the remaining 9 candidates yielded RFP-positive progeny after two rounds of injection (Table 14). In order to validate the targeted events that had occurred in the RFP-positive flies, genomic DNA was isolated and PCR was performed using primers flanking the targeted site. The PCR products were also purified and sequenced.

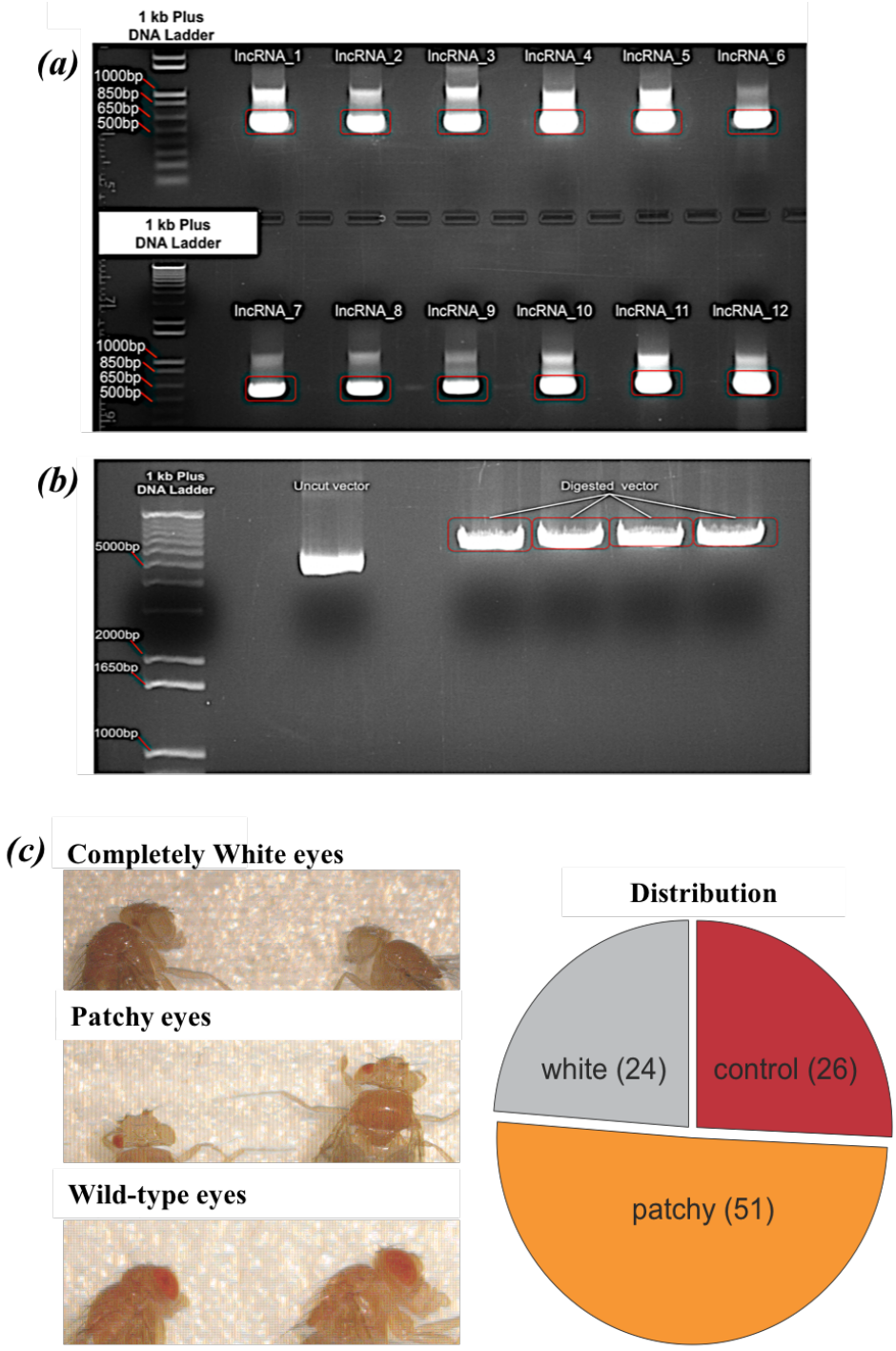


Figure 14: Somatic lncRNA transgenesis.
CRISPR-Cas9 catalyzed genome engineering using nonhomologous end joining (NHEJ) repair system. (a) Each pair of gRNAs were designed to target the promoter region of the lncRNA of interest. PCR products for lncRNA-1 to lncRNA-12. The expected size of the PCR fragment is ~600-nt. The bands corresponding to the desired size were excised extracted. (b) pCDF4 backbone was digested with BbsI-HF and dephosphorylated with CIP. The digested vector was size-selected by gel electrophoresis and a size of ~6,400bp was expected as shown. Uncut vector was used as a control. (c) CRISPR/Cas9 induced deletions of white gene. Transgenic lines expressing dual-gRNAs were generated by standard PhiC31-integrase–mediated transformation using injected DNA constructs. The dual gRNAs were expressed by U6:1and U6:3 promoters. Act5c-Cas9 expressing *Drosophila* was used to cross with dual-gRNA expressing flies that target the white gene as a pilot test. The first generation produces flies with mosaic patterns in both males and female. Approximately 50 % of the flies had mosaic expression patterns with small to large red patches in their eyes.

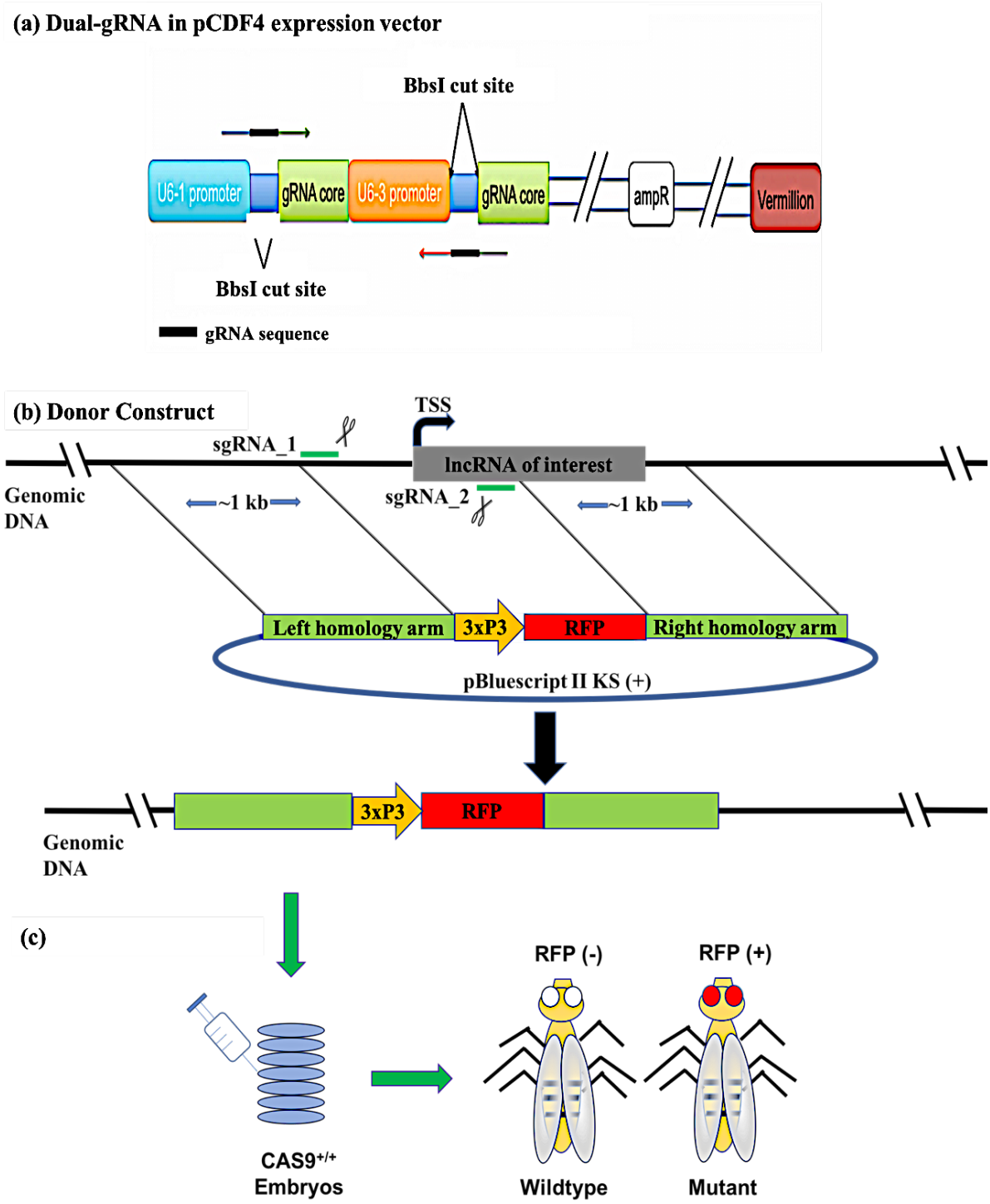

(a) Dual-gRNA in pCDF4 expression vector
BbsI cut site
U6-1 promoter
gRNA core
U6-3 promoter
gRNA core
ampR
Vermillion
BbsI cut site
gRNA sequence
(b) Donor Construct
TSS
sgRNA_1
lncRNA of interest
Genomic DNA
~1 kb
sgRNA_2
~1 kb
Left homology arm
3xP3
RFP
Right homology arm
pBluescript II KS (+)
3xP3
RFP
Genomic DNA
(c)
RFP (-)
RFP (+)
CAS9+/+ Embryos
Wildtype
Mutant

Figure 15: Schematic diagram of creating Transgenic CRISPR flies using HDR.
Genome editing with the use of homology-directed repair (HDR) mechanism to mend double-strand DNA breaks. This is a 2-parts strategy that requires a donor plasmid containing the repair template and the expression of gRNA from another DNA construct to be injected into embryos expressing Cas9. (a) The dual gRNA construct was generated using a forward and reverse primer sequence. Each oligonucleotide contains a gRNA sequence (black), a short sequence complementary to the 5' end of the *pCFD4* backbone as well as the 3' end of the backbone. The dual gRNA plasmids have been generated previously in section 3.2.3 and will be used as one of the two DNA plasmids that will be injected into embryos. (b) A donor construct was generated and composed of a RFP expression cassette that is driven by a 3xP3 promoter. A pair of specific homology arms adjacent to genomic site to remove are amplified from *Drosophila* genomic DNA and cloned to flank the RFP cassette. (c) The dual-gRNA construct was co-injected with the donor construct into Cas9-expressing flies. The donor construct is used as a template during HDR to replace the endogenous lncRNA locus with the RFP cassette. The use of HDR permits precise incorporation of the reporter gene at the genomic loci of the lncRNA and the expression of the fluorescence protein in the adult eye allows easy identification of successful knock-ins. Figure not to scale.

3.2.5 Test of survival to adulthood

To investigate the effects of the lncRNA mutation on the viability, heterozygous flies were crossed and their progeny were analysed for their ability to develop into an adult. The heterozygous mutant stocks possessed the lncRNA mutation over the balancer (balancer/mutant) and the crosses were maintained at normal temperature (25 °C). The total number for each genotype from the cross (balancer/mutant, mutant/mutant) were counted and the ratio of heterozygous/homozygous progenies was used to evaluate viability. From each cross, a 2:1 ratio of heterozygous to homozygous, lncRNA mutant flies is expected if the mutation is not lethal. However, there would be a change in the ratio if the knockout was lethal. The results of the crosses are shown in Table 15. LncRNA-3 and lncRNA-9 null mutants were found to be homozygous lethal. All of the progeny from these crosses were heterozygous mutants and none of them were homozygous for the lncRNA mutation, providing evidence that these two lncRNAs are essential for survival. For the other lncRNA mutants, I did not observe a change in the heterozygous/homozygous ratio, indicating that these lncRNAs are not essential for viability.

3.2.6 Validation of lncRNA mutagenesis

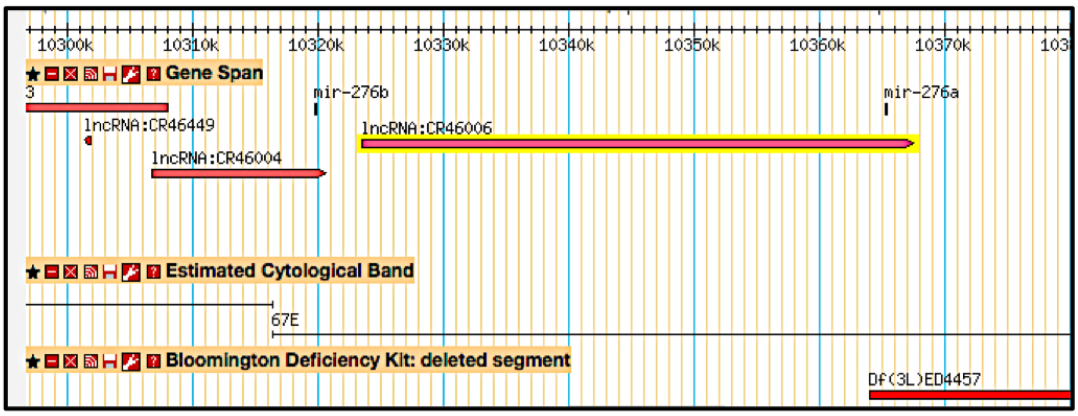
Validation with the use of deficiency line

For the lncRNA mutants with significant changes in the heterozygous/homozygous ratio, several methods were used to probe their viability. It has been shown in several studies that besides knockdown using siRNAs, CRISPR/Cas9 genome engineering was also an imperfect system and prone to off-target effects (Mali *et al.,* 2013, Cradick *et al.,* 2013, Schaefer *et al.,* 2017, Hsu *et al.,* 2013). These off-target mutations in other regions of the genome have been reported in other model organisms due to the binding specificity of each gRNA. Depending on the loci targeted and the design of the gRNA, the effect observed may vary from one gRNA to another. In this study, although the guide RNAs were selected with the highest stringency and sequences with no predicted off-target effects were chosen during the design phase, additional measures had to be taken to ensure that the potential off-targeting loci in lncRNA mutants were minimised and not affecting sequences in the genome that are influencing expression of nearby genes. The first strategy to verify the phenotype observed was to perform a genetic cross between the lncRNA heterozygous mutant and a deficiency (Df) line collection of *Drosophila* melanogaster that were generated with chromosomal

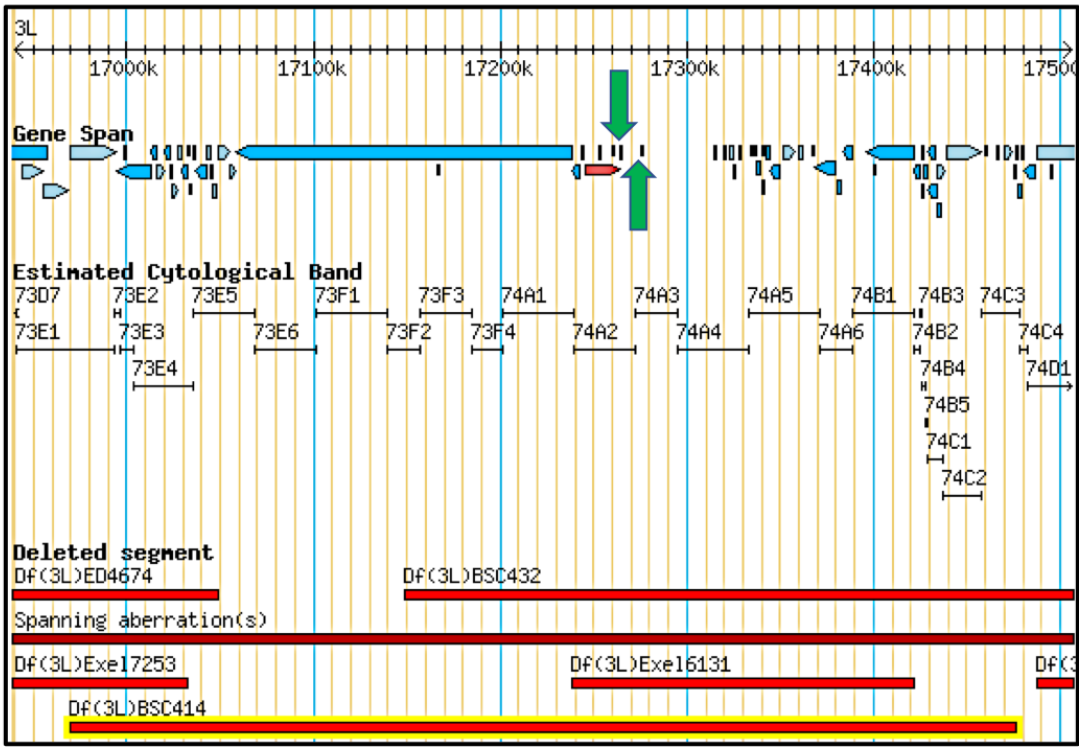
aberrations yet maintaining an isogenic background in the remainder of the genome (Ryder *et al.,* 2007). Most of these deletions are homozygous lethal and thus the deletions are kept in heterozygosity. The deficiency lines were obtained from Bloomington *Drosophila* Stock Centre and listed in Table 1. To unambiguously determine the link between the lethality and the lncRNA, the Bloomington deficiency line, Df(3L)ED4457 (BL9355) was crossed to lncRNA-3 heterozygous mutants, whereas Df(3L)BSC414 (BL24918) and Df(3L)BSC432 (BL24936) were used for the cross with lncRNA-9 heterozygous mutants (Figure 16a and Figure 16b respectively).

In case the lncRNA is required for viability, mutant/Df flies would not be observed in the next generation and the expected ratio would be skewed towards heterozygous flies. If the lethal phenotype observed in the homozygous CRISPR lines was due to an off-target effect, the mutant/Df flies would instead be expected to be viable as the two allele will complement each other. For lncRNA-9, both deficiency lines contained a deletion covering the lncRNA-9 genomic locus and only heterozygous flies were observed indicating that the deletion contained essential genes for survival (Figure 16b). In this case, I found that the none of the crosses produced KO/Df flies, indicating that the lncRNA was indeed required for survival. For lncRNA-3, Df(3L)ED4457 (BL9355) was the only fly stock available whereby the deleted segment in the deficiency line overlaps with miR-276a and only partially coincided with the 3' end of the lncRNA transcript. When the deficiency line was crossed to the lncRNA heterozygous siblings, flies that were Df/lncRNA-3-/- were found to be viable. The result suggests that the lncRNA-3 mutant phenotype observed is likely due to an off-target effect. However, the result has to be taken with caution as it might also highlight that the 5'end of the lncRNA transcript is sufficient for its function, with deletions of the 3'end tolerated. A similar cross was performed with a validated knockout strain for *mir-276a* (BL58906). In the F1 generation, flies with the Df/*mir-276a*$^{-/-}$ genotype was also found to be viable. The presence of viable Df/*mir-276a*$^{-/-}$ flies might also indicate that the annotation of this deficiency might be incorrect. Further investigation will be needed to validate this mutant and the deficiency line.

(a)


10300k
10310k
10320k
10330k
10340k
10350k
10360k
10370k
Gene Span
3
mir-276b
mir-276a
lncRNA:CR46449
lncRNA:CR46004
lncRNA:CR46006
Estimated Cytological Band
67E
Bloomington Deficiency Kit: deleted segment
Df(3L)ED4457

(b)


3L
17000k
17100k
17200k
17300k
17400k
Gene Span
Estimated Cytological Band
73D7
73E2
73E5
73F1
73F3
74A1
74A3
74A5
74B1
74B3
74C3
73E1
73E3
73E6
73F2
73F4
74A2
74A4
74A6
74B2
74C4
73E4
74B4
74D1
74B5
74C1
74C2
Deleted segment
Df(3L)ED4674
Df(3L)BSC432
Spanning aberration(s)
Df(3L)Exel7253
Df(3L)Exel6131
Df(3L)BSC414

(c)

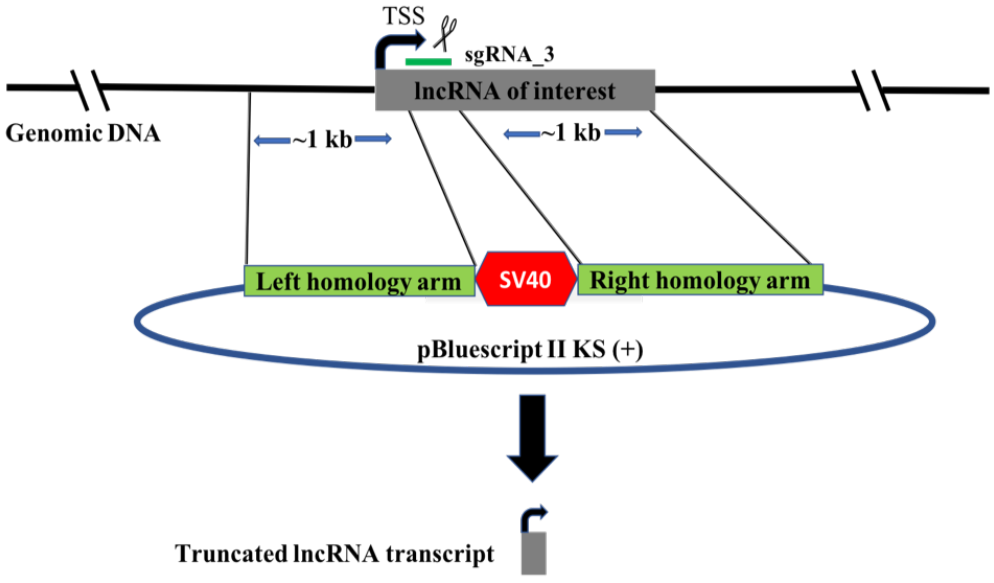

TSS
sgRNA_3
lncRNA of interest
Genomic DNA
~1 kb
~1 kb
Left homology arm
SV40
Right homology arm
pBluescript II KS (+)
Truncated lncRNA transcript

Figure 16: Validation of lncRNA mutagenesis
(a,b) Genome browser view of lncRNA-3 and lncRNA-9. (a) Df(3L)ED4457 (BL9355) is a deficiency line found on the third chromosome. The cytological bands 67E2 through 68A7 shows the position of the deficiency strains used for complementation tests with lncRNA-3. This deficiency line only overlaps partially at the 3' end with lncRNA transcript. (b) As lncRNA-9 was unannotated in the genome browser, the arrows (green) were used to indicate the approximate location of the lncRNA. The lncRNA is located between CG13723 and lncRNA:CR45436 (indicated by green arrow). Df(3L)BSC414 (BL24918) is a deficiency line that encompasses a deletion from cytological band 73E1 to 74C3 whereas Df(3L)BSC432 (BL24936) includes a deletion from cytological band 73F2 to 74E4. Both deficiency lines overlap with the lncRNA locus and were used to validate the lncRNA phenotype. (c) Insertion of SV40 transcriptional stop signal into lncRNA loci using Cas9-HDR. A donor construct was generated and contained a SV40 transcriptional stop signal sequence that is flanked by a pair of specific homology arms that are cloned from *Drosophila* genomic DNA. The gRNA construct was co-injected with the donor construct into Cas9-expressing flies and used as a template during HDR replacing the endogenous lncRNA locus with the SV40 stop signal. With the introduction of an early polyA signals close to the TSS to terminate transcription, a presumably non-functional lncRNA product will be formed.

LOF mutation following integration of transcriptional termination signal

A parallel approach to prove the specificity of the observed phenotype was undertaken by inserting a transcriptional stop signal (SV40) downstream of the transcriptional start site (TSS) for lncRNA-3 and lncRNA-9 using CRISPR/Cas9. We argue that this method would give us an independent mutation that was different from the knock-in and it was also improbable for the same off-target effects to be observed using different gRNAs. Details of the cloning strategy are shown in Figure 15c. Instead of a dual-gRNA strategy, I opted for a single cut strategy with the use of a gRNA to target the genomic loci of the lncRNA. The aim of this approach is to insert a transcriptional stop signal cassette into the cut site. The gRNA target sequence (20-nt) was again selected using online tools (http://www.flyrnai.org/crispr/) and cloned into the pCDF3 expression vector. For each lncRNA candidate, 2 gRNAs that target different regions of the lncRNA transcript were designed to be close to the TSS and more than 1000 nucleotides away from any annotated gene. Similarly, the protocol and expression vector used (pCDF3) used were described for *Drosophila* earlier (Port *et al*., 2014). The pCDF3 vector containing a *U6-3* promoter that drives the expression of the gRNA. The protocol used for the Cas9-HDR design in paragraph 3.2.4 was taken to clone the donor vector. The final constructs were verified via sanger sequencing prior to fly injection (Figure 15c).

The pool containing a single gRNA with one donor was injected into Cas9 expressing embryos under the control of vasa regulatory sequences (BL55821: y[1] M{vas-Cas9.RFP-}ZH-2A w[1118]/FM7a, P{w[+mC] = Tb[1]}FM7-A). Due to the low efficiency of the Cas9 mutagenesis, I requested the injection of ~400 embryos (equivalent to 2 microinjection) with each pool set. As the donor plasmids do not contain any selection marker, the founders had to be screened by PCR and sanger sequencing. For the first round of microinjection, no positive insertion was detected from the PCR screen that was carried out on 60-80 viable founders from each pool. A second round of microinjection (~400 embryos) was performed with re-purified plasmids and the founders were screen for positive insertion using the same method. Out of the 2 gRNAs that were used to target lncRNA-9, positive founders were found from one of them (pES103a, pES107c) and the fly stock were balanced (Table 16). A similar viability assay in section 3.2.5 was performed with the two lncRNA-9 mutant lines using heterozygous mutants and the null mutants were found to be homozygous lethal, confirming the earlier results observed with the CRISPR HDR mutant line for

lncRNA-9 (Table 15). For lncRNA-3, also the second round of microinjection failed to generate positive founders for both gRNAs (Table 16).

3.2.7 LncRNA-9 is essential for the completion of embryogenesis

The assay probing the survival to adulthood identified two lncRNAs, lncRNA-3 and lncRNA-9, as required for *Drosophila* development, with null mutants for these lncRNA candidates not surviving to the adult stage. To investigate at which stage of development these mutants perish, the embryos (wildtype, sibling heterozygous mutant and homozygous mutant) were plated in a 6 x 48 well plates containing cornmeal and allowed to develop at 25 °C for at least 36 hr. For each lncRNA candidate, the heterozygous were sorted from the homozygous mutants by hand using the YFP signal from fluorescence balancer. From this screen, I found that both heterozygous and homozygous null mutant embryos for lncRNA-3 were viable as they emerge into first instar larvae and phenotypically indistinguishable from the control wild-type embryos. This suggesting that despite the fact that lncRNA-3 is expressed at embryonic stages, it is not essential for embryonic development which can still complete successfully in its absence. On the other hand, the majority of the lncRNA-9 homozygous null mutant embryos failed to hatch (Figure 17a). A close examination of these embryos under a brightfield microscope revealed that some of them were still viable as they appeared to be moving their mouthpiece within the membrane. A very small proportion of the lncRNA-9 homozygous null mutant embryos hatched into first instar larvae but their growth arrested and they died before the second instar stage (Supplementary figure 1).

3.2.8 Homozygous lncRNA-3 null mutants were embryonic viable but die during pupation

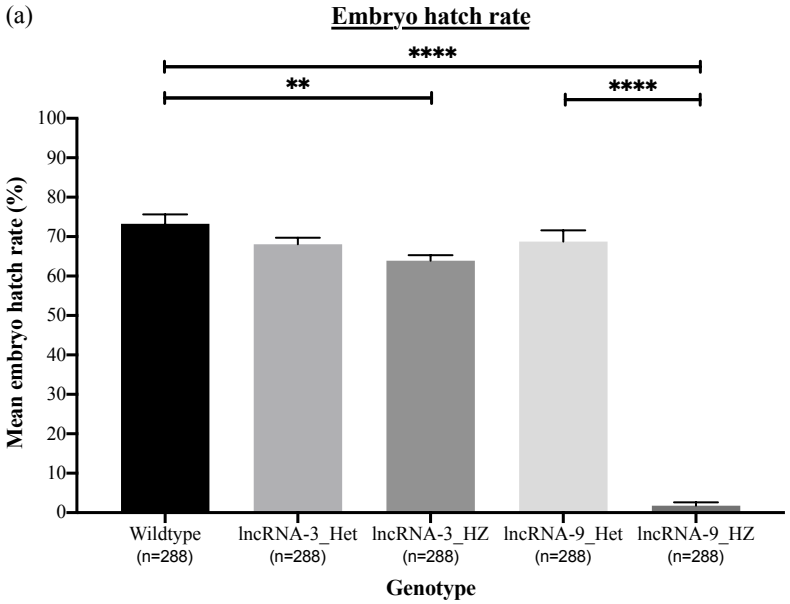
In order to decipher the *in vivo* consequences of lncRNA-3 loss of function, the larvae obtained from the embryogenesis screen (section 3.2.7) were transferred to fresh vials and incubated at 25 °C. The heterozygotes displayed similar viability as wildtype control larvae and completed metamorphosis as they continued to develop through larval and pupal stages (Figure 17b). On the other hand, the lncRNA-3 homozygous null embryos survived embryogenesis and hatched into first instar larvae. They are able to complete all stages of larval development (Figure 17b). However, they arrest during pupation and eventually developed into pharate adults without emerging from the puparium (Figure 17c). A separate gross examination of the homozygous null mutants revealed that during the larval-pupal

transition, the larvae appeared to have developed to different extents compared to wildtype and heterozygous siblings (Supplementary figure 2). Compared to wildtype and heterozygous larvae, it took the 3rd instar larvae of the homozygous mutant more than 24 hrs to begin crawling up the side of the vial and enter the pupal stage. The vials containing homozygous mutants were allowed to developed for a total of 30 days from the time that the embryos were collected and checked every five days. I observed that the mutants were at different stages of pupal development and at the end of 30 days, all of them were arrested at the pharate adult stage. The pharate adults were subsequently removed from their puparium and examined under a dissection microscope. In contrast to a control flies, the head of the pharate adults were located much more anteriorly than normal and was unable to extend fully (Supplementary figure 3).

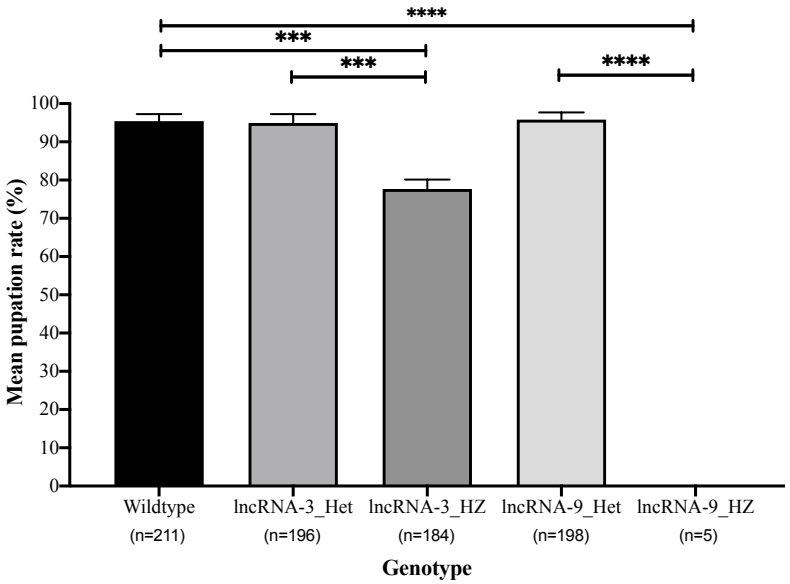
3.2.9 Expression of *miR-276a* is independent of lncRNA-3

To further investigate the biological function of lncRNA-3, we looked at its genomic locus and found that the transcript for the lncRNA candidate did not overlap with any annotated coding and noncoding genes when the RAMPAGE data was mapped to the *D. melanogaster* genome assembly (dm3, BDGP Release 5). However, following lift-over of the lncRNA to the latest genome assembly (dm6, BDGP Release 6 plus ISO1 MT), five different isoforms were identified for lncRNA-3. Surprisingly, all of the isoforms for lncRNA-3 overlapped with the *Drosophila* microRNA miR-276a. This miRNA was initially found to be located >37Kb downstream of lncRNA-3 in the previous genome assembly. Previous studies have shown that a recessive lethal mutation of miR-276a is associated with *Drosophila* learning behaviour. To investigate whether the mutation of the lncRNA-3 transcript had a detrimental effect on the expression of miR-276a, we performed a complementation cross between the heterozygous mutants for the lncRNA and miR-276a (BL58906: w*; TI{TI}mir-276aKO/TM3, P{GAL4-twi.G}2.3, P{UAS-2xEGFP}AH2.3, Sb1 Ser1). The lncRNA deletion complemented the lethality and confirmed that decreased viability observed in our lncRNA-3 mutant could not be ascribed to the loss of miR-276a activity. This observation excludes that our mutation strategy could have inadvertently generated the loss of essential DNA sequences driving miR-276a expression and it also confirms that the expression of miR-276a is independent of that of lncRNA-3 (Figure 18).

(a)


Embryo hatch rate
****
**
****
Mean embryo hatch rate (%)
100
90
80
70
60
50
40
30
20
10
0
Wildtype (n=288)
lncRNA-3_Het (n=288)
lncRNA-3_HZ (n=288)
lncRNA-9_Het (n=288)
lncRNA-9_HZ (n=288)
Genotype

(b)


Pupation rate
****
***
***
****
Mean pupation rate (%)
100
90
80
70
60
50
40
30
20
10
0
Wildtype (n=211)
lncRNA-3_Het (n=196)
lncRNA-3_HZ (n=184)
lncRNA-9_Het (n=198)
lncRNA-9_HZ (n=5)
Genotype

(c)

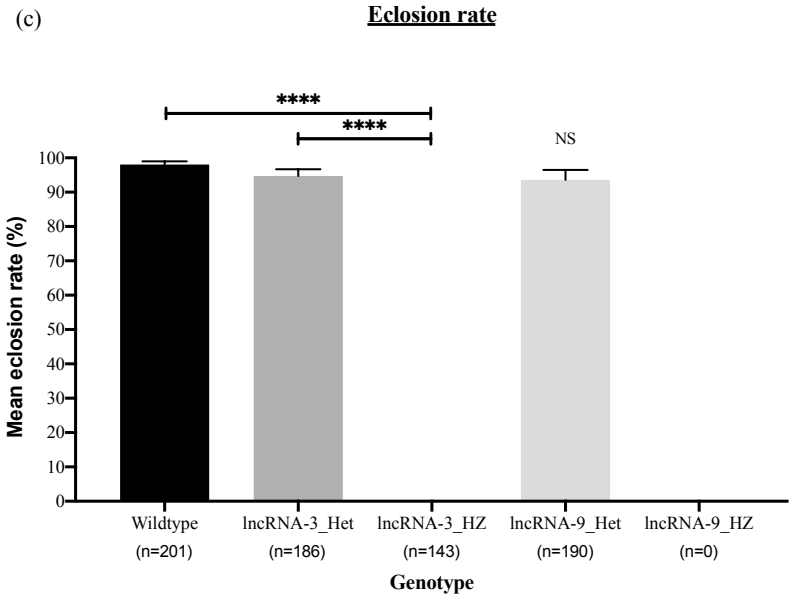

Eclosion rate
****
****
NS
Mean eclosion rate (%)
100
90
80
70
60
50
40
30
20
10
0
Wildtype (n=201)
lncRNA-3_Het (n=186)
lncRNA-3_HZ (n=143)
lncRNA-9_Het (n=190)
lncRNA-9_HZ (n=0)
Genotype

Figure 17: Analysis of lncRNA mutants at different stages of development

(a) Mutants of lncRNA-9 display reduced embryo viability. In the screen for completion of embryogenesis, the frequency of egg-hatching was used as a measurement of embryo viability. The embryos were collected and plated in 48-well plates containing apple juice agar and incubated at 25 °C for further development. A total of six data points were collected (n=288). The number of larvae that have emerged were counted and plotted in percentage as a function of survival relative to wildtype controls. (b) LncRNA-3 is not necessary for larval development. To extent the analysis, the larvae that hatched were allowed to continue developing in fresh vials containing cornmeal at 25 °C. The number of third instar larvae that have emerged from the food and have undergone pupation were counted and plotted in percentage as a function of survival relative to wildtype. Although majority of lncRNA-3 mutants undergoes pupation, a significant increase in mortality is observed in homozygotes during the larval-pupal transition when compared to both controls and heterozygotes. In addition, the homozygous mutant larvae for lncRNA-3 completed larval development at a much slower rate compared to the wildtype and heterozygous mutants. For lncRNA-9 homozygous null mutants, the escapees arrested at the first instar larval stage and did not complete the larval life cycle. (c) LncRNA-3 mutant is necessary for eclosion. The pupae that have completed pupation and emerged into adult fruit flies were counted and plotted similarly to the earlier stages of development. Flies carrying homozygous mutation for lncRNA-3 were found to develop to pharate adults but did not emerge from the eclosion. Homozygous null mutants for lncRNA-9 were omitted from the analysis as the mutants were mainly embryonic/larval lethal. A comparison of the wildtype with the lncRNA-9 heterozygous siblings did not display any statistical difference in development following pupation. Bar graphs represent mean values and error bars are ± SEM. The statistical significance was determined with the Student's t-test.

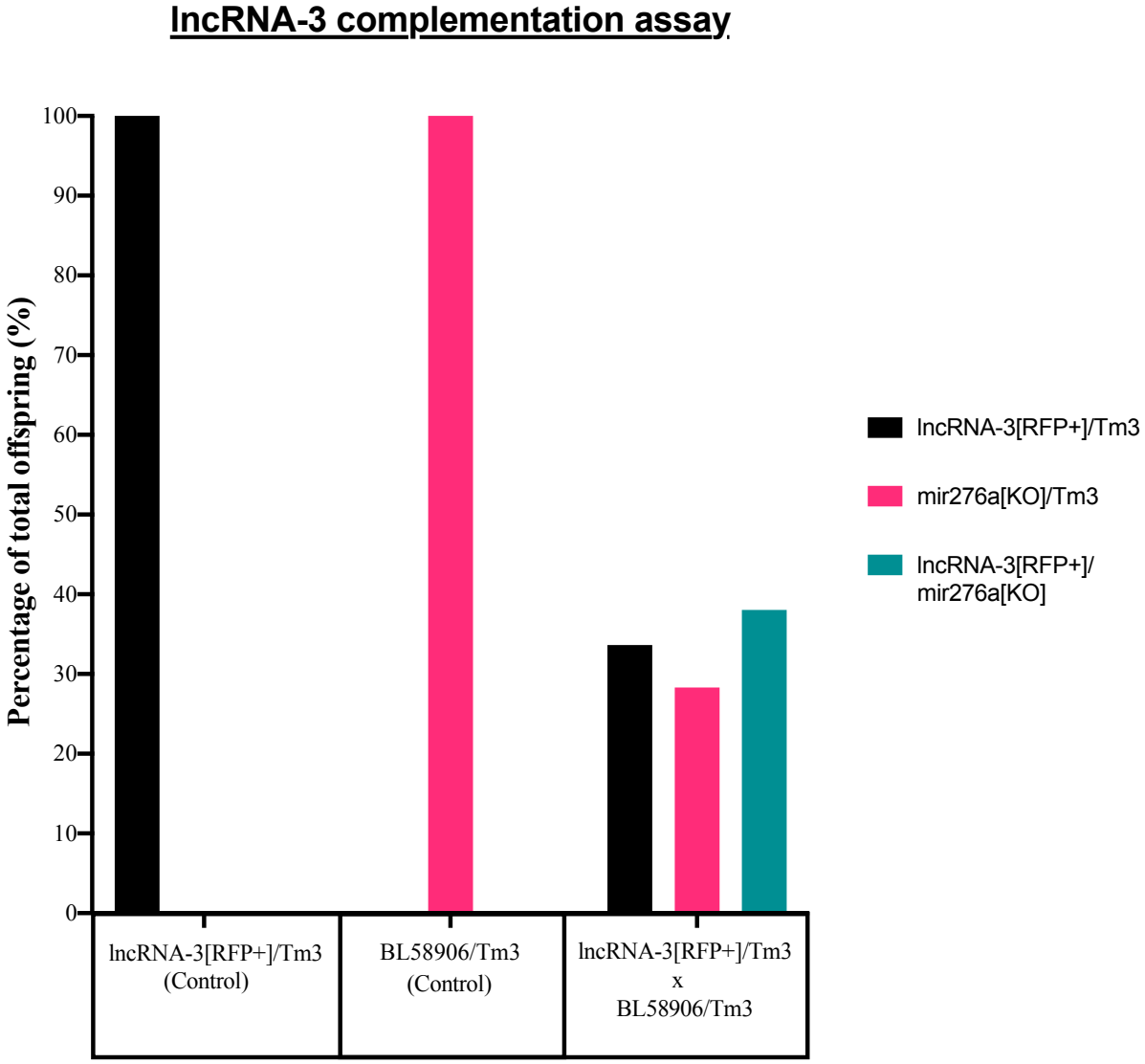


Figure 18: Expression of lncRNA-3 is independent of *Drosophila miR-276a*.
miRNA-276a overlaps with the 3' end of lncRNA-3. A genetic cross was carried out to determine if the progeny carrying the lncRNA-3 mutant allele and the miR-276a[KO] allele were viable. The parental strains were used as control. The number of adult offspring was counted and each column represents the fraction of the total number of progenies. For both lncRNA and miRNA single mutants, only heterozygous siblings were viable. However, the lncRNA-3/miR-276a double mutants were viable and developed into adults.

3.2.10 Male and female fertility test

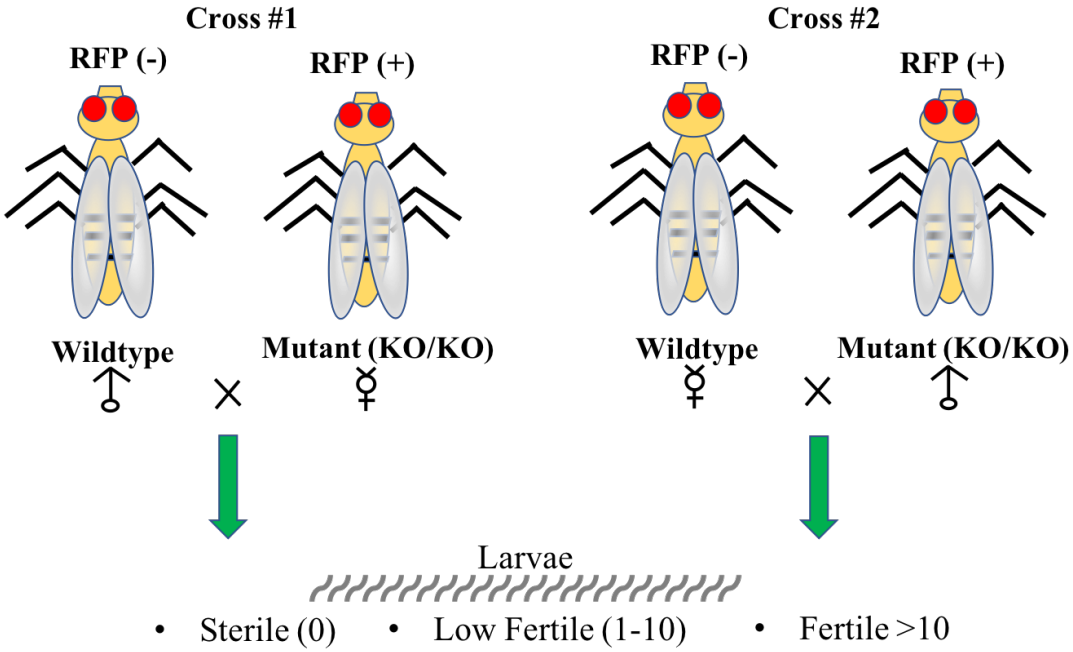
LncRNAs have previously been reported to be required for the fertility of an organisms and to be important for proper spermatogenesis and ovary development in *Drosophila* (Wen *et al.,* 2016; Nyberg *et al.,* 2016). To investigate if any of the lncRNAs under analysis had a role in fertility, genetic crosses were setup using homozygous viable lines and wildtype flies. The experiment was performed using a protocol described earlier (Chen *et al*., 2014). To set up the crosses, freshly hatched wildtype males and null mutant virgin females were collected in separate vials and mated after 3 days (Figure 19). A reciprocal cross was also set up as shown in Figure 19. Each vial was visually examined to ensure eggs had been laid after 5 days and the parents were discarded. The total number of larvae was recorded. A male/female will be considered to be sterile if there were no larvae hatched in any vial, and fertile if there were more than 10 larvae in each vial. The mutant would be considered to show reduced fertility in the event that there were between 1 to 10 larvae in each vial. LncRNA-3 and lncRNA-9 were excluded from this experiment as they do not generate homozygous mutant adults. Overall, I did not observe sterility for any of the novel lncRNA null mutants in either of the reciprocal crosses. I have also dissected the ovaries of all homozygous lncRNA mutants and did not observe different morphology compared to wildtype ovaries (Figure 19). Therefore, I concluded that none of the eleven lncRNAs under analysis has a role in fertility in either male or female flies.

3.2.11 Simple larvae locomotion assay

LncRNAs have also been reported to contribute to cognition and behavioural changes (Li *et al.,* 2012). To investigate if any of the candidates had a locomotion defect, 50 null mutants from each lncRNA candidate were transferred onto a larva crawling assay. The assay setup is shown in Figure 20a. Yeast paste was placed behind the finishing line and used as a food attractant for the larvae. The larvae were arranged behind the starting line and the number of larvae that crossed the finishing line was scored at the end of the test (after 30 min). I found no significant differences between the tested lncRNA null mutants and the wildtype flies (Figure 20b). I have also collected ten lncRNA-9 homozygous null first instar larvae and assayed them in this setup. None of the larvae moved beyond the starting line, thus they appear to have a locomotion defect. Due to the small number of mutants used and the fact that these larvae are likely to have additional defects, however, the data for lncRNA-9 was not included in the data analysis for this test.

3.2.12 Gross anatomical examination of the adult fruit fly

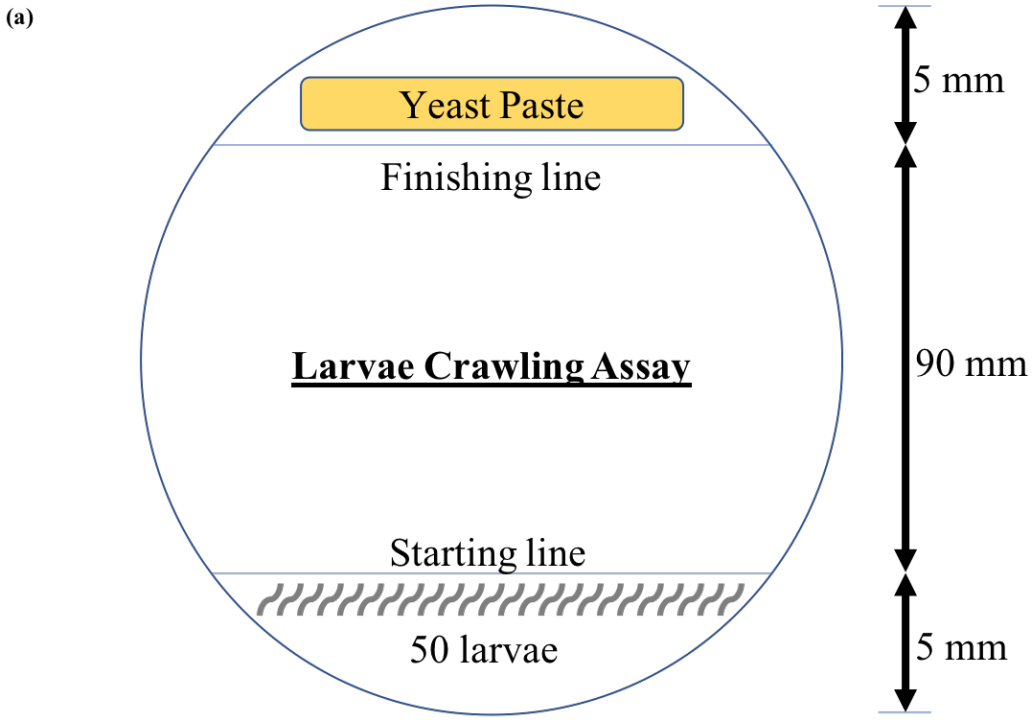
To investigate if any of the lncRNA null mutants had any anatomical abnormalities, I performed a gross examination of the external morphology of the lncRNA mutant adult flies using six qualities (a pair of wings, a pair of halteres, 3 pair of legs, a head, thorax and the eight abdominal segments). Wildtype flies were used as a reference point when the comparisons were made. For each lncRNA candidate, six heterozygotes and six homozygous null mutants were examined carefully under a dissecting microscope for any possible developmental defects (three males and females each). Except for lncRNA-6 (*bxd*) homozygous null mutants, I did not observe anatomical differences in any of the lncRNA mutants. For *bxd* lncRNA, it was previously reported that the pair of halteres is replaced by a second pair of wings in the *bxd* mutant (Lewis, 1978; Bassett *et al.,* 2014). Here, besides a partial transformation of the halteres in homozygous mutant adults, I did not find any other abnormalities (Supplementary Figures 4-16).



Fertility	Cross #1	Cross #2
	Female	Male
	KO/KO	KO/KO
lncRNA-3	N.A	N.A
lncRNA-4	Fertile	Fertile
lncRNA-5	Fertile	Fertile
lncRNA-6	Low	Low
lncRNA-9	N.A	N.A
lncRNA-10	Fertile	Fertile
lncRNA-11	Fertile	Fertile
lncRNA-12	Fertile	Fertile
lncRNA-13	Fertile	Fertile
lncRNA-14	Fertile	Fertile
lncRNA-15	Fertile	Fertile
lncRNA-16	Fertile	Fertile
lncRNA-19	Fertile	Fertile

Figure 19: Individual lncRNA have no overt function in fertility

A fertility screen was set up using a reciprocal genetic cross between wildtype flies and homozygous null lncRNA mutants. In this experimental setup, ten crosses were performed with each vial containing a single male and virgin female fly (10 replicates). Cross #1 was set up using a homozygous null virgin female with a wildtype male. Cross #2 was performed using a homozygous null male with a wildtype virgin female to determine if either male or female mutant are sterile or showed reduced fertility. For both crosses, the parents were discarded after five days and the total number of larvae that hatched from the vials were counted. For each vial, the male or female mutant were considered to be sterile if no larvae were found, low fertility if there were between 1-10 larvae and fertile if there were more than ten larvae observed in each vial. From this screen, the majority of lncRNA mutant flies were fertile and an average of 12 larvae were counted in each vial. For *bxd* lncRNA (lncRNA-6), the number of embryos that hatched were between 1-10 and thus it was considered to be lowly fertile. No significant difference was observed between lncRNA homozygous mutants and wildtype for both t-test and one-way ANOVA test.


(a)
Yeast Paste
Finishing line
Larvae Crawling Assay
Starting line
50 larvae
5 mm
90 mm
5 mm

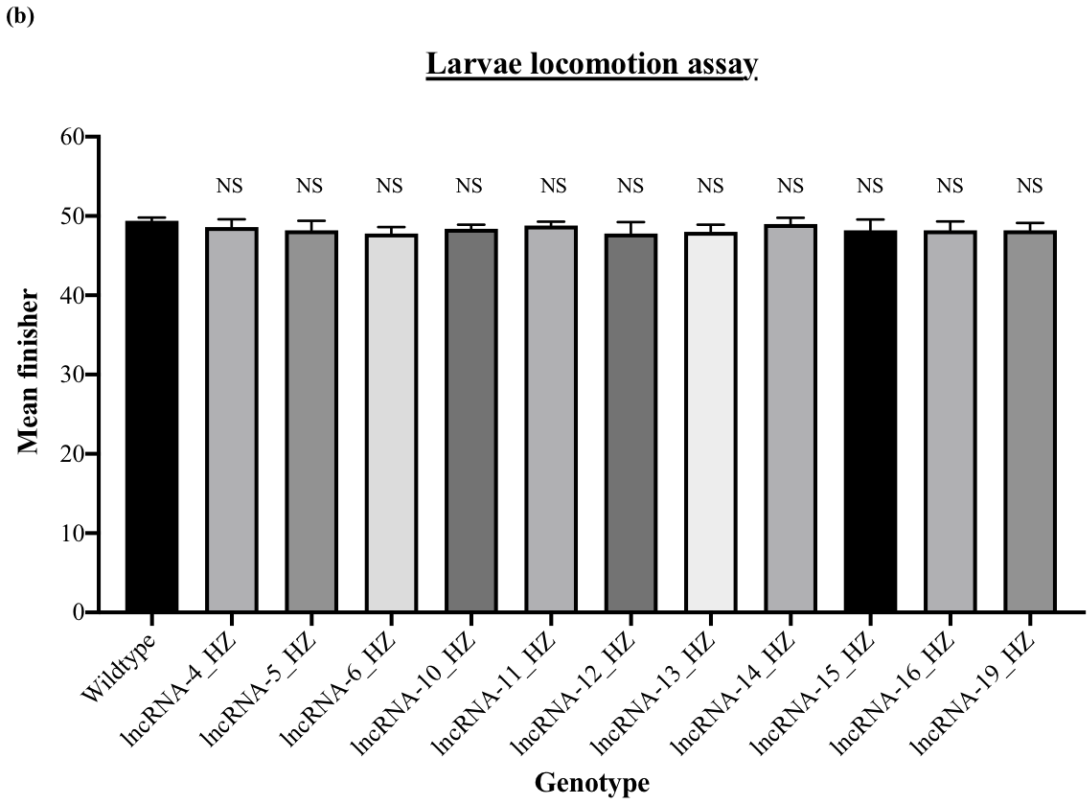

(b)
Larvae locomotion assay
Mean finisher
60
50
40
30
20
10
0
NS
Wildtype
lncRNA-4_HZ
lncRNA-5_HZ
lncRNA-6_HZ
lncRNA-10_HZ
lncRNA-11_HZ
lncRNA-12_HZ
lncRNA-13_HZ
lncRNA-14_HZ
lncRNA-15_HZ
lncRNA-16_HZ
lncRNA-19_HZ
Genotype

Figure 20: LncRNA mutation did not affect larvae locomotion.
(a) Schematic representation of the larval crawling assay performed on an apple juice agarose plate. (b) The larval crawling assay of wild-type (WT) and the various homozygous lncRNA mutants generated. The lncRNA candidates (lncRNA-3 and lncRNA-9) with a homozygous lethal phenotype were excluded from this assay. Larvae (n = 50 each per datapoint) were plated behind the starting line and allowed to crawl to toward the yeast paste without any interference. The number of larvae that crosses the finishing line were scored after 30 min (Not to scale). For each genotype, a total of 5 data points were collected. During each new set of experiment, a fresh agar plate containing fresh yeast paste was used. No significant difference was observed between lncRNA homozygous mutants and wildtype larvae (for both t-test and one-way ANOVA test). Error bars represent ± SEM.

Table 11: LncRNA Candidates and their Coordinates (dm3)

lncRNA-	ID	TSC Coordinates (dm3)	Strand	Size (kb)	Annotation (dm3)	CPAT Coding Probability (dm3)	Upstream gene name	Size (Kb)	Distance to lncRNA (Kb)	Downstream gene name	Size (Kb)	Distance to lncRNA (Kb)
1	FBgn0051781_01	chr2L:16869345-16869508	+	2.58	CR31781	4.57E-02	CG6012	1.35	8.07	CG32832	0.68	0.30
2	FBgn0264479_01	chr3L:11686221-11686254	+	0.53	CR43887	1.3E-01	CG14135	1.11	0.38	CG11652	3.13	0.19
3	FBgn0265415_02	chr3L:10319728-10319881	+	9.40	-	6.92E-07	*Or67d*	1.42	52.00	CG12362	1.96	86.19
4	FBgn0050009_02	chr2R:6151308-6151344	-	2.50	CR30009-RA	8.36E-04	CG12911	4.59	1.32	CG12910	2.34	3.03
5	FBgn0019661_05	chrX:3756065-3756103	-	2.45	*roX1*-RA	5.53E-04	*ec*	45.68	1.27	*yin*	5.36	0.54
6	FBgn0020556_01	chr3R:12598828-12598999	-	1.26	*Bxd*-RC	1.86E-02	*Ubx*	78.25	9.29	*Glut3*	1.71	15.12
7	FBgn0263380_02	chr3L:14970662-14970908	-	0.69	-	3.38E-02	*Ocho*	1.19	0.59	CG3349	2.24	2.48
8	FBgn0047095_01	chrX:19068891-19068949	-	0.55	pncr004:X-RA	1.70E-02	*Inx5*	1.89	0.44	CG7556	2.98	2.10
9	RAMPAGEpeak_002227	chr3L:17262841-17262866	+	0.70	-	6.11E-08	CG13723	0.33	4.91	CG6485	0.85	50.22
10	RAMPAGEpeak_005280	chr3L:13106986-13107015	-	0.77	-	7.97E-03	CG11251	1.39	23.90	*trn*	3.82	0.39
11	RAMPAGEpeak_006440	chr2L:20484386-20484497	+	0.90	-	4.98E-03	CG2493	1.87	20.94	CG34007	0.34	10.61
12	RAMPAGEpeak_008667	chr2L:17585024-17585143	-	0.37	-	1.11E-01	CG7094	1.41	29.26	*Oli*	1.76	4.48
13	RAMPAGEpeak_011922	chrX:7160282-7160322	-	0.73	-	1.53E-01	CG9650	13.97	25.83	CG1958	0.98	0.66

LncRNA-	ID	TSC_Coordinates (dm3)	Strand	Size (kb)	Annotation (dm3)	CPAT Coding Probability (dm3)	Upstream gene name	Size (Kb)	Distance to lncRNA (Kb)	Downstream gene name	Size (Kb)	Distance to lncRNA (Kb)
14	RAMPAGEpeak_015813	chr2R:5437743-5437834	+	0.83	-	2.38E-02	CG1888	1.78	3.28	CG12931	1.47	9.36
15	RAMPAGEpeak_016800	chr2R:9250322-9250340	-	1.01	-	5.47E-03	CG17048	0.53	8.49	CG10814	1.74	13.15
16	RAMPAGEpeak_018802	chr3R:22519840-22519855	+	0.59	-	2.68E-02	*TwdlQ*	0.89	0.60	*amon*	16.46	3.58
17	RAMPAGEpeak_024168	chr3R:22443183-22443186	-	0.76	-	2.87E-03	CG5467	6.39	13.44	*TwdlM*	1.11	1.66
18	FBgn0264857_02	chr3R:12748808-12749095	-	2.63	*iab-8*	5.01E-03	*Abd-A*	22.41	1.08	*Abd-B*	6.90	4.59
19	FBgn0264857_05	chr3R:12674246-12674550	-	1.72	*iab-8*	1.49E-02	*Abd-A*	22.41	1.08	*Abd-B*	6.90	78.85
20	FBgn0262972_01	chr3R:7081835-7081859	-	3.48	CR43283	5.01E-03	CG5214	2.58	1.24	KP78b	2.08	90.25
21	RAMPAGEpeak_001315	chr3L:10299762-10299782	+	7.94	-	1.02E-05	*Or67d*	1.42	32.04	CG12362	1.96	106.15
22	RAMPAGEpeak_021570	chr3R:8070463-8070515	+	0.36	-	1.73E-02	*Cyp313a3*	1.83	5.63	CG3942	2.41	9.45

The information in this table relate to the mapping of the lncRNAs on the dm3 *Drosophila* genome. The first column indicates the lncRNA number, the second indicates the name of the RAMPAGE peak for unknown lncRNAs or the ID for already annotated lncRNAs. The column highlighted in blue indicates the direction of transcription in the genome. The neighbouring protein-coding genes closest to the lncRNA are indicated and their direction of expression relative to the lncRNA is indicated in either green (sense to the lncRNA) or yellow (anti-sense to the lncRNA).

Table 12: LncRNA Candidates and their Coordinates (dm6)

LncRNA-	ID	TSS Coordinates (dm6)	*CNCI score	Flybase Annotation	Remarks
1	FBgn0051781_01	chr2L:16869345-16869508	-0.0937984	CR31781	-
2	FBgn0264479_01	chr3L:11693121-11693154	-0.0958464	CR43887	-
3	FBgn0265415_02	chr3L:10326628-10326781	-0.0225280	CR46006	Overlaps with miR-276a
4	FBgn0050009_02	chr2R:10263803-10263839	-0.0004096	CR30009	-
5	FBgn0019661_05	chrX:3862032-3862070	-0.0393216	roX1	-
6	FBgn0020556_01	chr3R:16773106-16773277	-0.0532480	bxd	Overlaps with CG31275
7	FBgn0263380_02	chr3L:14977562-14977808	-0.0303104	CR43432	-
8	FBgn0047095_01	chrX:19174858-19174916	-	CG33939	Protein-coding gene
9	RAMPAGEpeak_002227	chr3L:17269741-17269766	-0.0860160	-	May overlap with CR45436
10	RAMPAGEpeak_005280	chr3L:13113886-13113915	-0.0532480	-	-
11	RAMPAGEpeak_006440	chr2L:20484386-20484497	-0.0536576	CR45361	Overlaps with miR-1 and CR44980
12	RAMPAGEpeak_008667	chr2L:17585024-17585143	-0.0176128	CR45112	-
13	RAMPAGEpeak_011922	chrX:7266249-7266289	undetermined	Intronic region of CR44357?	-

LncRNA-	ID	TSS Coordinates (dm6)	*CNCI score	Flybase Annotation	Remarks
14	RAMPAGEpeak_015813	chr2R:9550238-9550329	undetermined	-	-
15	RAMPAGEpeak_016800	chr2R:13362817-13362835	-0.1187840	CR45312	-
16	RAMPAGEpeak_018802	chr3R:26694118-26694133	-0.0245760	-	-
17	RAMPAGEpeak_024168	chr3R:26617461-26617464	undetermined	-	
18	FBgn0264857_02	chr3R:16923086-16923373	-0.0253952	iab-8	Overlaps with iab-4 and CR43617
19	FBgn0264857_05	chr3R:16848524-16848828	-0.0004096	iab-8	-
20	FBgn0262972_01	chr3R:11256113-11256137	-0.0823296	CR43283	-
21	RAMPAGEpeak_001315	chr3L:10306662-10306682	-0.0159744	CR46004	-
22	RAMPAGEpeak_021570	chr3R:12244741-12244793	-0.0004096	CR45582	-

*Transcripts with a CNCI score of less than 0 would be considered as non-coding. Due to lack of transcript structural information, three of the lncRNA candidates were not identified and the scores were underdetermined using CNCI.

The information in this table refer to the lncRNA lift-over from dm3 to the dm6 *Drosophila* genome. LncRNA-8 was highlighted in grey as it was annotated as a protein-coding gene in dm6 and it will therefore be removed from this study. Newly annotated lncRNAs in dm6 were highlighted in green.

Table 13: Stage- and tissue-specific information of lncRNA candidates

LncRNA-	ID	Embryo	Larvae	Pupae	Head	Digestive system	Testes	Ovaries
1	FBgn0051781_01	1	1	1	1	1	1	0
2	FBgn0264479_01	1	0	0	0	0	0	0
3	FBgn0265415_02	1	1	1	1	0	0	0
4	FBgn0050009_02	1	1	1	1	0	0	0
5	FBgn0019661_05	1	1	1	1	1	1	0
6	FBgn0020556_01	1	1	1	0	0	0	0
7	FBgn0263380_02	1	1	1	1	1	0	0
8	FBgn0047095_01	-	-	-	-	-	-	-
9	RAMPAGEpeak_002227	1	1	1	1	0	0	0
10	RAMPAGEpeak_005280	1	1	1	1	0	0	0
11	RAMPAGEpeak_006440	1	1	1	1	0	0	0
12	RAMPAGEpeak_008667	1	0	0	0	0	0	0
13	RAMPAGEpeak_011922	1	0	0	0	0	0	0

LncRNA-	ID	Embryo	Larvae	Pupae	Head	Digestive system	Testes	Ovaries
14	RAMPAGEpeak_015813	1	1	1	1	0	0	0
15	RAMPAGEpeak_016800	1	1	1	1	0	0	0
16	RAMPAGEpeak_018802	1	0	0	0	0	0	0
17	RAMPAGEpeak_024168	1	0	0	0	0	0	0
18	FBgn0264857_02	1	1	0	0	0	0	0
19	FBgn0264857_05	1	1	0	0	0	0	0
20	FBgn0262972_01	1	1	1	0	0	0	0
21	RAMPAGEpeak_001315	1	1	1	1	0	0	0
22	RAMPAGEpeak_021570	1	1	1	1	0	0	0

This table contains information on the expression of the lncRNA during development according to CAGE reads tracks in Dreos *et al.,* 2013. For any tissue or stage, the number "1" was used to denote expression and "0" to indicate an absence of the transcript.

*LncRNA-8 was highlighted in grey as it was annotated as a protein-coding gene in dm6 and will be removed from this study.

Table 14: CRISPR microinjection results (TSS replacement with 3xP3-RFP cassette)

LncRNA-	ID	Dual-gRNA plasmid #	Donor plasmid #	Deletion type	Deletion Size (Kb)	Cas9 strain used	# of founders / injection cycle		Overall Success rate
							1st	2nd	
1	FBgn0051781_01	pES_1	pES_31	TSS-del	1.67	BL55821	0	0	0 %
2	FBgn0264479_01	pES_2	pES_32	TSS-del	0.63	BL55821	0	0	0 %
3	FBgn0265415_02	pES_3	pES_33	TSS-del	1.76	BL55821	0	3	< 1 %
4	FBgn0050009_02	pES_4	pES_34	TSS-del	0.24	BL55821	0	3	< 1 %
5	FBgn0019661_05	pES_5	pES_35	TSS-del	1.00	TH_attP2	3	-	< 1 %
6	FBgn0020556_01	pES_6	pES_36	TSS-del	0.96	BL55821	2	-	< 1 %
7	FBgn0263380_02	pES_7	pES_37	TSS-del	1.25	BL55821	0	0	0 %
8	FBgn0047095_01	pES_8	pES_38	TSS-del	1.12	TH_attP2	0	0	0 %
9	RAMPAGEpeak_002227	pES_9	pES_39	TSS-del	1.00	BL55821	0	1	< 1 %
10	RAMPAGEpeak_005280	pES_10	pES_40	TSS-del	1.59	BL55821	0	3	< 1 %
11	RAMPAGEpeak_006440	pES_11	pES_41	TSS-del	2.04	BL55821	4	-	< 1 %
12	RAMPAGEpeak_008667	pES_12	pES_42	TSS-del	0.92	BL55821	4	-	< 1 %

LncRNA-	ID	Dual-gRNA plasmid	Donor plasmid	Deletion type	Deletion Size (Kb)	Cas9 strain used	# of founders / injection cycle		Overall Success rate
							1st	2nd	
13	RAMPAGEpeak_011922	pES_13	pES_43	TSS-del	1.28	TH_attP2	0	3	< 1 %
14	RAMPAGEpeak_015813	pES_14	pES_44	TSS-del	0.87	BL55821	4	-	< 1 %
15	RAMPAGEpeak_016800	pES_15	pES_45	TSS-del	0.79	BL55821	3	-	< 1 %
16	RAMPAGEpeak_018802	pES_16	pES_46	TSS-del	0.86	BL55821	0	3	< 1 %
17	RAMPAGEpeak_024168	pES_17	pES_47	TSS-del	1.35	BL55821	0	0	0 %
18	FBgn0264857_02	pES_18	pES_48	TSS-del	2.54	BL55821	0	0	0 %
19	FBgn0264857_05	pES_19	pES_49	TSS-del	2.00	BL55821	0	2	< 1 %
20	FBgn0262972_01	pES_20	pES_50	TSS-del	1.03	BL55821	0	0	0 %
21	RAMPAGEpeak_001315	pES_21	pES_51	TSS-del	2.16	BL55821	0	0	0 %
22	RAMPAGEpeak_021570	pES_22	pES_52	TSS-del	2.53	BL55821	0	0	0 %

The rows highlighted in grey indicate that no positive founders were generated from the CRISPR microinjection.

Table 15: *Drosophila melanogaster* viability test results

Viability Assessment of KO lines				
Name	Line #	Heterozygous Progeny	Homozygous Progeny	Heterozygous/ Homozygous Ratio
Stocks with lncRNA TSS removed and replaced with 3xP3-RFP cassette				
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	2M	166	0	3:0
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	3M	148	0	3:0
w; +; lncRNA-3 [RFP+]/Tm3-YFP*	6M	161	0	3:0
w; lncRNA-4 [RFP+]/CyO-YFP; +*	1M	127	72	~2:1
w; lncRNA-4 [RFP+]/CyO-YFP; +*	3M	127	77	~2:1
w; lncRNA-4 [RFP+]/CyO-YFP; +*	5M	111	51	~2:1
lncRNA-5 [RFP+]/Fm7-YFP; +; +	1M	169	95	~2:1
lncRNA-5 [RFP+]/Fm7-YFP; +; +	2M	174	80	~2:1
lncRNA-5[RFP+]/Fm7-YFP; +; +	3M	175	88	~2:1
w; +; lncRNA-6 [RFP+]/Tm3-YFP*	2M	96	23	~2:1
w; +; lncRNA-6 [RFP+]/Tm3-YFP*	4M	87	26	~2:1
w; +; lncRNA-9 [RFP+]/Tm3-YFP*	1M	168	0	3:0
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	1M	125	86	~2:1
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	3M	140	87	~2:1
w; +; lncRNA-10 [RFP+]/Tm3-YFP*	5M	158	90	~2:1
w; lncRNA-11 [RFP+]/CyO-YFP; +*	1M	144	66	~2:1
w; lncRNA-11 [RFP+]/CyO-YFP; +*	2M	161	49	~2:1
w; lncRNA-11 [RFP+]/CyO-YFP; +*	3M	125	61	~2:1
w; lncRNA-11 [RFP+]/CyO-YFP; +*	4M	150	71	~2:1
w; lncRNA-12 [RFP+]/CyO-YFP; +*	1M	111	48	~2:1
w; lncRNA-12 [RFP+]/CyO-YFP; +*	2M	160	79	~2:1
w; lncRNA-12 [RFP+]/CyO-YFP; +*	3M	135	55	~2:1
w; lncRNA-12 [RFP+]/CyO-YFP; +*	4M	112	53	~2:1
lncRNA-13 [RFP+]/Fm7a-YFP; +; +	1M	164	87	~2:1
lncRNA-13 [RFP+]/Fm7a-YFP; +; +	3M	157	54	~2:1
lncRNA-13 [RFP+]/Fm7a-YFP; +; +	5M	128	40	~2:1
w; lncRNA-14 [RFP+]/CyO-YFP; +*	1M	169	95	~2:1
w; lncRNA-14 [RFP+]/CyO-YFP; +*	2M	137	87	~2:1
w; lncRNA-14 [RFP+]/CyO-YFP; +*	3M	157	65	~2:1
w; lncRNA-14 [RFP+]/CyO-YFP; +*	4M	134	57	~2:1
w; lncRNA-15 [RFP+]/CyO-YFP; +*	1M	145	65	~2:1
w; lncRNA-15 [RFP+]/CyO-YFP; +*	2M	141	86	~2:1
w; lncRNA-15 [RFP+]/CyO-YFP; +*	3M	133	66	~2:1
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	2M	166	79	~2:1
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	3M	180	86	~2:1
w; +; lncRNA-16 [RFP+]/Tm3-YFP*	4M	162	77	~2:1
w; +; lncRNA-19 [RFP+]/Tm3-YFP*	1M	144	70	~2:1
w; +; lncRNA-19 [RFP+]/Tm3-YFP*	2M	158	65	~2:1
Stocks with SV40 insertion downstream of lncRNA TSS				
w; +; lncRNA-9 [SV40]/Tm3-YFP*	2M	205	0	3:0
w; +; lncRNA-9 [SV40]/Tm3-YFP*	3M	199	0	3:0

The viability test was carried out to check for lncRNAs that were essential for *Drosophila* development. For each genetic cross, ten heterozygous adults from each line (five males and females each) were placed in a vial and transferred to a new vial for two more consecutive times. The progenies from all three vials were scored based on their genotype. Visible markers from the balancer chromosome was used to distinguish the progenies. For the mutants that carry the CyO balancer, flies with indistinguishable wings or broken wings were discarded from count. The rows highlighted in grey indicates the lncRNAs for which no homozygous RFP-positive progenies were observed. Note that for lncRNA-9 both HDR and SV40 mutant lines were generated.

Table 16: CRISPR microinjection results (SV40 insertion)

ID	Dual-gRNA plasmid #	Donor plasmid #	Distance from TSS (bp)	Cas9 strain used	# of founders / injection cycle		Overall Success rate
					1st	2nd	
lncRNA-3	pES_102a	pES106c	898	BL55821	0	0	0 %
	pES_102b	pES106d	80	BL55821	0	0	0 %
lncRNA-9	pES_103a	pES107c	187	BL55821	0	10	< 1 %
	pES_103b	pES_107d	410	BL55821	0	0	0 %

The rows highlighted in grey indicate that no positive founders were generated from the CRISPR microinjection. Out of the ten viable founders for lncRNA-9, two lines were selected and balanced after verification by PCR and sequencing.

3.3 Discussion

The extensive use of next generation sequencing technologies has shown that tens of thousands of genomic sites are pervasively transcribed, producing several classes of ncRNAs such as miRNAs, endogenous siRNAs, piRNAs and lncRNAs (Derrien *et al.,* 2012; Wen *et al.,* 2014, Brown *et al.,* 2014). In recent years, we have seen an explosion in the discovery and characterization of a class of biologically significant RNA transcripts that are now known as the lncRNAs (Ponting *et al.,* 2009; Guttman *et al.,* 2009; Hon *et al.,* 2017). Compared to protein-coding genes, lncRNAs were found to be under less stringent evolutionary constraints in terms of their primary sequence, and they are generally expressed at lower levels (Derrien *et al.,* 2012). To gain a better understanding of the biological functions of lncRNA, I have chosen *Drosophila* as a model organism given its long history in genetic studies and the large number of genetic tools available. I have decided to focus on development and particularly embryogenesis because of the availability of convenient datasets to mine. We posited that problems during this very well studied stage would likely result in recognisable and hopefully interpretable phenotypes. It is important to note that for our purpose it is not really critical whether the lncRNA expression is only restricted to embryonic development or shows broader expression.

Using RAMPAGE data generated previously (Batut *et al.,* 2014; Batut *et al.,* 2017), we explored the expression profile of lncRNAs in five species (*Drosophila melanogaster, Drosophila simulans, Drosophila erecta, Drosophila ananassae,* and *Drosophila pseudoobscura)*. Out of the 25,426 TSCs identified in *Drosophila melanogaster*, ~14% of them could not be associated with any known gene or sRNA based on paired-end RAMPAGE analysis and were therefore identified as putative lncRNA candidates. With the addition of 291 TSCs that have been previously associated with Flybase-annotated lncRNA genes, a list containing 3,973 lncRNA TSCs was produced. The previous analysis of these RAMPAGE data already suggested a potential biological role for lncRNAs during embryonic development as evidenced by the strong conservation and sequence constraints that the promoters for some of these lncRNAs are under. Based on conservation of both promoter sequence and its activity across species, we have first narrowed down this list to 631 lncRNA TSCs that are expressed in all five *Drosophila* species. This is a more rigorous list of candidates compared to previous methods of annotation (Young *et al.,* 2012) because it includes a much more comprehensive RNA-Seq dataset collected at hourly time-points during embryogenesis from across five species. We then decided to focus on a small group of candidate lncRNAs among the ones

that showed the best conservation also in terms their expression profile. We further narrowed down the selection to focus on lncRNAs with high expression in our model system, *Drosophila melanogaster*, and that were clearly distinct from neighbouring annotations as ideal candidates for further experimental investigation into their contributions to developmental processes. In the end, we obtained a list of 20 candidate lncRNAs plus 2 control lncRNAs (lncRNA-6/*bxd* and lncRNA5/*rox1*) that I brought forward for CRISPR mutagenesis.

Although the existing RAMPAGE libraries focused on samples collected during embryogenesis, the RAMPAGE method can theoretically be extended in the future to encompass other stages of development or tissue types. A comparison of the activity of each individual lncRNA TSC in all 5 species at other developmental stages would allow to further gauge the functional conservation of these lncRNA loci throughout the *Drosophila* life cycle. From the analysis of available modENCODE CAGE read data, we infer that some of the lncRNA candidates we chose are also expressed at other stages of development. As already pointed out, we choose for our analysis lncRNAs that showed expression during embryonic stages but we did not require for the lncRNAs not to be expressed in any other stage and our analysis of their potential function is not affected if a lncRNA is also expressed in other stages of development. Recently, several studies have proposed that functional lncRNAs are only activated at a specific stage of development (Sweta *et al.,* 2019; Bhatia *et al.,* 2019; Young *et al.,* 2012) and this could be observed for five out of 21 of the lncRNA candidates analysed. On the other hand, lncRNAs can also be found to be transcribed at several developmental stages as seen for the bulk of the lncRNA candidates investigated. This is not surprising as *Drosophila* undergoes three main stages of growth that encompass several moulting cycles during larval development which are followed by a pupation stage before the final molt to the adult stage. The upregulation of lncRNAs at these respective stages could signify that they play an essential role for transformation and organogenesis and deserves further exploration (Chen *et al.,* 2016).

In order to understand the physiological functions of *Drosophila* lncRNAs, it is crucial to remove these transcripts in vivo and examine the phenotypic consequences of these mutations (Mattick, 2013). Previous studies investigating the functions of lncRNAs have indicated that these transcripts have roles in different developmental processes and diseases (Wu *et al.,* 2016; Ran *et al.,* 2016). Our early understanding of lncRNA function in the regulation of gene expression were mainly derived from loss-of-function studies performed in vitro using a

myriad of animal cells lines (Ma *et al.,* 2016; Iyer *et al.,* 2015). Several lncRNAs, including *MALAT1* and *NEAT1* were found to have a function in diseases such as cancer (Serviss *et al.,* 2016). However, a disparity was observed between the phenotypes that were found in vitro versus those that were observed in vivo. Both *MALAT1* and *NEAT1* were found to be important regulators of cellular processes. However, mice with deletions that removed either lncRNAs had no detectable phenotypes (Zhang *et al.,* 2012; Eißmann *et al.,* 2012). These confounding results thus highlighted the need for loss-of-function studies in vivo as one of the gold standards when studying lncRNA functions.

3.3.1 CRISPR/Cas9-mediated gene silencing of lncRNA transcripts

Presently, phenotypic and functional information were provided for the majority of protein-coding genes identified through genetic screens. However, mutational analysis of non-coding regions of the genome has largely been lacking. The identification of lncRNA transcripts has for a long time relied on computational predictions (Young *et al.,* 2012; Kellis *et al.,* 2014). However, to understand the functionality of lncRNAs, experimental validation is required. Follow-up experimental validation of the list of candidates has been the primary aim after comprehensive annotations from the RAMPAGE data. Beyond computational predictions of functionality from publicly available data, I have assessed the in vivo phenotypic effect of thirteen lncRNA knocking-outs. CRISPR/Cas9-mediated genome engineering is an efficient and rapid technique to create novel, targeted mutations in a chosen locus. In this study, I have used *Drosophila* as the model organism to test the application of HDR-mediated genomic editing of lncRNA transcripts using CRISPR/Cas9-based systems. Out of the 21 lncRNA candidates selected, I obtained an overall success rate of 62 %, with 13 mutants generated that included two positive controls.

I first tested to create *Drosophila* lncRNA mutants using a “somatic CRISPR” approach (Port *et al.,* 2014). I have adopted the approach of having the dual-gRNA expressing flies and Cas9 expressing flies in a separate entity so as to avoid creating mutants with lethal phenotypes. Flies carrying dual-gRNA constructs targeting various lncRNAs were successfully established. I was able to observe a subset of offspring with mutations in the white gene, which served as a pilot test for a mutation with an observable phenotype. However, a large proportion of the G_1 population presented mosaic patterns in the eyes. The mutation efficiency was not 100 % and required tools for rapid identification of targeted events. Two issues surfaced immediately. The first issue was genotyping. Without visible

marker, PCR-based molecular screening was the only option to screen for successful CRISPR knockout events, but it is time consuming and labour intensive to generate and maintain candidate fly stocks. The second issue was heritability of the mutant phenotype. Most of the CRISPR-based studies in *Drosophila* were performed in annotated protein-coding genes and mutants were backcrossed to check for inheritance of mutant genes. However, this was challenging in our case as the study focused on un-annotated lncRNAs with unknown functions.

Most approaches that have been used to create CRISPR mutants were heavily dependent on readily identifiable phenotypes for identification of successful targeting events (Bassett *et al.,* 2013; Gratz *et al.,* 2013a; Port *et al.,* 2014). The functions of these phenotypes were known prior to the study, which allowed for quick identification of mutant flies. In order to rectify the two issues mentioned above, I have taken a second approach to create lncRNA mutants using CRISPR/Cas9-mediated HDR with dsDNA donors. This is a more universally applicable method that allows screening for targeting events using visible markers. A pair of flanking homology arms and the visible marker 3xP3-RFP were incorporated into a donor vector. Using this technique, I was able to develop transgenic lines and also screen for successful targeting events without the need for large-scale genotyping by PCR to identify founders with the desired insertion.

Nevertheless, the task of creating donor constructs was more challenging than previously imagined. The first issue was the quality and concentration of genomic DNA used. For the extraction of genomic DNA of higher purity (ratio of OD 260/280 is between 1.8 -1.9) from *Drosophila*, a commercial kit was used. The second problem I faced was the designing of primers to amplify each homology arm as it was difficult to amplify from some of the regions of genomic DNA such as those that were rich in GC sequences. The donor construct was initially assembled using a 4-way (pBS II vector backbone, a pair of homology arms, 3xP3-RFP fragment) Hifi-DNA Gibson assembly mix. However, this is not efficient for assembling four fragments together. Therefore, I have made use of overlapping PCRs to join the inserts together before assembling them with the vector backbone. This resulted in a higher success rate overall linking two fragments together instead of joining four fragments.

The Cas9/HDR efficiency had been reported to be generally low even in cell-based systems and it was not surprising that we had a <1 % efficiency with the majority of the candidates, requiring at least 400 embryos to be injected for each lncRNA locus (Liu *et al.,*

2012; Yu *et al.,* 2014; Bassett *et al.,* 2014). To improve the efficiency of HDR when combining with CRISPR-Cas9, others have shown that by inhibiting components of the NHEJ pathway one could bias the repair system toward HDR (Bier *et al.,* 2018; Beumer *et al.,* 2008). However, a higher efficiency was not always observed in flies and further research into cellular repair mechanisms during development would help us to develop better tools for manipulating the genome (Gratz *et al.,* 2014; Ge *et al.,* 2016). The study of lncRNA function using different genome editing methods have always been met with different methodological caveats. They can for example cause disruption of regulatory elements surrounding the genomic loci. Therefore, the use of additional tools such as RNAi would represent a useful complementary approach allowing to perturb the transcript without disturbing the locus itself (Guttman *et al.,* 2011). RNAi would be more effective in suppressing cytoplasmic lncRNAs and the use of CRISPR interference (CRISPRi) would allow the investigation of nuclear lncRNA functions (Lennox and Behlke, 2016; Liu *et al.,* 2017). Besides generating mutations in the lncRNA, little is known about the over- or mis-expression of these transcripts. The beauty of working with *Drosophila* was the availabilty of genetic tools that have been generated. It would be interesting to express the candidate lncRNAs in a tissue or temporal specific manner using different Gal4 lines and observe for other phenotypic changes.

Challenges of the Cas9-HDR methodology

The induction of targeted mutations in *Drosophila* has become possible through the advancement of genome editing tools such as CRISPR to induce mutations through DNA double-strand breaks followed by DNA repair with non-homologous end joining (NHEJ) or homology-directed repair (HDR) mechanisms. In this study, I have made used of CRISPR/Cas9 with HDR to replace the promoter region of the lncRNA with a reporter gene. Of the 21 lncRNA, 13 lncRNA mutants were successfully generated whereas no founders were established for the other eight candidates. Furthermore, compared to non-lethal mutations, little is known about the HDR recovery rates for recessive lethal alleles. Several factors such as the choice of injected *Drosophila* strain or the selection of gRNA could have contributed to the lower efficiency observed.

One of the possible issues that led to the low number of positive offspring could be attributed to the preference of the cellular repair mechanism used. To compared the efficiency of CRISPR/Cas9 genome engineering, a cocktail of dsDNA construct, gRNA plasmids and Cas9 proteins were introduced in *lig4* mutants and a higher targeting proficiency was observed (Gratz *et al.,* 2014, Yu *et al.,* 2014). However, more validation will be required for

the use of *lig4* mutants to enhance HDR finding as contrasting results have been reported (Ge *et al.,* 2016; Ren *et al.,* 2014).

The second important parameter is the prevalence of polymorphisms between distinct genetic backgrounds in *Drosophila*. The sequences used for the production of gRNAs and homology arms of the donor construct are generally taken from the reference genome assembly and may differ from the sequences in the injected strain. Mismatches between gRNAs and DNA sequences have been reported to decrease the overall the binding affinity and cleavage by Cas9 (Jinek *et al.,* 2012, Cong *et al.,* 2013, Mali *et al.,* 2013). Similarly, polymorphism close to the cleavage site may also influence the binding of the DNA donor template with the genomic DNA and thus contributes to a lower HDR rate (Gratz *et al.,* 2015). Therefore, sequencing of the target genomic locus in the genetic background in which the genome engineering experiment will be performed should ideally be performed as a single base pair change at the intended target site could have detrimental effect to the success of the experiment. Moreover, the majority of the cleavage efficiency scores have been obtained from high throughput experiments in S2 cells to predict the cleavage efficiency of a given gRNA (www.flyrnai.org/evaluateCrispr/). It is however not fully understood if similar efficacy would be observed in vivo.

The choice of gRNAs plays an important role as it has been shown that the mutagenesis efficiency is positively correlated to the number of GC content present at the six nucleotides that are closest to the PAM sequence (Ren *et al.,* 2014; Bassett *et al.,* 2013, Kondo and Ueda, 2013, Yu *et al.,* 2013). Those gRNAs with low GC content close to the PAM sequence tend to produce sub-optimal mutagenesis efficiency rates due to the unstable gRNA:DNA heteroduplex in Cas9 (Jinek *et al.,* 2014, Nishimasu *et al.,* 2014). However, little is known regarding the association between efficiency and overall GC content and further experiments should be carried out. With the newer developments in algorithms for the selection of gRNAs, higher HDR rates can be obtained by careful selection of target sites (Heigwer *et al.,* 2016; Port *et al.,* 2019).

From a procedural point of view, the overall concentration of the cocktail containing the gRNA plasmid and donor DNA template injected have been shown to influence the success rate of the experiment. In this study, the ratio of DNA template and dual-gRNA construct used was 2:1. Several studies have demonstrated that optimisation of plasmid concentration is needed for HDR-directed repair as it greatly increases the viability of the flies

(Ren *et al.,* 2014, Port *et al.,* 2014). Having a high concentration of gRNAs tend to induce an adverse effect on the embryos which results in the decreased number of progenies (Ren *et al.,* 2014). Therefore, to efficiently generate mutagenesis, an equal concentration of DNA donor template and gRNA construct in the range of 50 to 100 ng/µl was suggested (Ren *et al.,* 2014).

As a whole, there were several factors that could have contributed to the low recovery rate seen in our experiment and it is difficult to pinpoint the exact reason. With the advancement of the CRISPR/Cas9 technology and algorithms for the selection of gRNAs, several improvements can be undertaken to improve the experimental design so as to enhance the overall efficiency of the CRISPR-Cas9 methodology and at the same time increase the number of offspring obtained with successful mutagenesis.

3.3.2 Systematic analysis of lncRNA functions

To date, many questions were raised about the relevance of these long molecules and what their functions in various biological processes are as only a small fraction of lncRNAs has been explored mechanistically, especially in *Drosophila* (Young *et al.,* 2012; Matthews *et al*., 2015). Therefore, the next step forward for this project was to characterise the mutants generated. A simple systematic functional study of the mutants for the 13 lncRNAs was used to address their significance in *Drosophila melanogaster*. The pipeline was broken down into three parts; namely viability and morphogenesis, fertility, and locomotion.

Completion of *Drosophila* development

The first part of the analysis was focused on understanding if these lncRNAs were required for survival and development as these are phenotypes that were identifiable under the microscope. The mutants for lncRNA-9 and lncRNA-3 were found to be essential for survival. The homozygous null mutants for these lncRNAs were unable to complete the full developmental cycle and no adult survivors were recovered. To further examine the authenticity of these lethal mutations, screens of deletion mutations, commonly known as deficiencies (Df), were used. Both KO/KO and KO/Df combinations for lncRNA-9 further confirmed that the lncRNA mutation phenotype observed was not due to an off-target effect of CRISPR/Cas9. In addition, similar to the validation of RNAi approaches, it is common practice to authenticate the results with a second independent gRNA construct. A second loss-

of-function experiment was performed by inserting transcriptional terminators into the genomic loci of both lncRNAs. A strong polyadenylation signal was inserted close to the 5′ end of the lncRNA so as to arrest transcription and result in a likely non-functional transcript (Bond *et al.,* 2009; Grote *et al.,* 2013). Using this method, homozygous null mutants for lncRNA-9 displayed a similar lethal phenotype to that observed with the promoter deletion. In all instances, the lncRNA transcript itself was shown to be essential and not the DNA of the lncRNA locus.

Due to the lack of cytologically mapped deficiencies that overlap the region that was removed or the entire genomic region of lncRNA-3, the phenotype was first validated using a deletion that overlapped with *miR-276a* and the 3' end of the lncRNA. Surprisingly, both lncRNA-3^{KO}/Df and Df/*miR-276a*^{KO} progenies were viable and thus suggestive of a possible off-target effect. However, this result has to be carefully interpreted. Given that the lncRNA mutant was generated by removing the promoter region and the 5' end of the transcript, there is a likelihood that the 3' end of the lncRNA transcript is not essential for its biological function. The Df line should be sequenced to ensure that the deleted region really corresponds to the one annotated in databases and actually overlaps with mir-276. The fact that lncRNA-3 mutants can complement *mir-276a* (BL58906) clearly indicates that the observed lncRNA-3 phenotype does not depend on *mir-276a* and that lncRNA-3 does not influence the expression of this miRNA. Secondly, I have attempted to generate loss-of-function mutants with transcriptional terminators as a complementary approach but was unsuccessful in obtaining any founders. The use of several gRNAs to specifically alter the genetic code at defined positions in the lncRNA genome loci is still the recommended approach to minimise the risk of off-targets and will be the method to further confirm the specificity to the lncRNA-3 mutant phenotype.

The regulation of Hox gene expression during animal development have been extensively studied in *Drosophila melanogaster*. Most Hox genes are important for the allocation of distinct morphological identities to each body segment of the organism. With the recent advances in technology, a myriad of lncRNA sequences embedded within the Hox gene clusters and their flanking regions have been discovered (e.g. *bxd, lncRNA:iab-4, lncRNA:iab-8*). These new players provide important regulatory inputs to both temporally and spatially restrict the patterns of Hox gene expression and mutations of these ncRNAs could possibly lead to homeotic transformations. To determine if the loss of lncRNA candidates results in developmental defects, I have compared the body segments of each homozygous

null lncRNA mutant to the wildtype adult. In this study, I have considered the adult fruit fly as composed of six sections (head, wings, legs, abdominal segments, haltere and thorax region). In all cases, I did not observe any morphological abnormalities in the homozygous null mutants under analysis except for lncRNA-3. Careful dissection of the puparium of the lncRNA-3 mutant revealed traces of developmental defects in the head region which resembles a mild form of head eversion during pupal ecdysis. This observation may hint towards a potential role of lncRNA-3 in head or brain development. Subsequent layers of investigation will be required to understand how lncRNA-3 contributes to this phenotype and the signalling pathways that are affected by its loss.

Following the assessment of lncRNAs that were required for survival to adulthood, I have identified lncRNA-3 and lncRNA-9 as both required for development. From the tissue-specific CAGE data, I was able to confirm that both lncRNAs are expressed during *Drosophila* development with peaks of expression spanning from embryogenesis to pupation. Furthermore, the expression of both lncRNA-3 and lncRNA-9 was detected in the adult head of the fruit fly. This observation made me question at which stage the lethality ensues. By performing a survivability test at three distinct stages of development (embryo-larvae, larvae-pupae, pupae-adult), I was able to show that the majority of the lncRNA-9 homozygous null mutants were embryonic lethal with a significant reduction in embryonic hatching rate. Although there was a small number of escapees, they quickly arrested as 1st instar larvae with no animals developing into final instar larvae, pupae, or adults. LncRNA-3 homozygous null mutants, on the other hand, completed embryogenesis and hatched into first instar larvae. Our findings for larval and pupal development confirmed that lncRNA-3 homozygous null mutants were pupal lethal with the mutant flies arresting at late pupal stages. A significant increase in mortality rate was also observed during the larval and pupal transition. While the homozygous null mutants for lncRNA-3 were found to complete larval development as seen in wildtype embryos, they displayed unsynchronised and late larval-pupa transition, suggesting either a potential developmental delay in these mutants. A microRNA downstream of the TSS was found to overlap with lncRNA-3 and previous studies have shown that miR-276a mutants were also pupal lethal (Li *et al.,* 2013, Chen *et al.,* 2014). Subsequent follow-up complementation assay indicated that the expression of lncRNA-3 has no effects on *miR-276a* expression.

Other developmental analyses

In this part of the analysis, homozygous null mutants were used and lncRNA-3 and lncRNA-9 were omitted from this study because viable mutants could not be obtained. The second part of the study was to analyse if lncRNAs were involved in reproduction and germ cell development. Recent studies have shown that lncRNAs were highly expressed in the testis and contributed to the reproductive fitness of the animal (Wen *et al.,* 2016; Kurihara *et al.,* 2017). We reasoned that defects in lncRNAs required for germ cell development would allow overall development to take place but cause infertility. The fertility of both male and females were measured and none of the homozygous null mutants were sterile. The result was not surprising as a further analysis of the tissue expression of these candidates showed that none of the lncRNAs tested were expressed in the testis or ovaries. The last part of the pipeline aims to characterise lncRNAs that may be involved in neurological disorders. An increasing number of studies have shown that lncRNAs play roles in the development of the nervous system and are expressed in specific cells of the central or peripheral nervous system (Graveley *et al.,* 2011; Ponjavic *et al.,* 2009; Chalei *et al.,* 2014; Raveendra *et al.,* 2018). Furthermore, lncRNAs (e.g. *Drosophila bereft* and *yar*) might be involved in regulating neural development and cognitive processes. Over the years, different behavioural assays were generated in *Drosophila*. Examples include larval locomotion, adult negative geotaxis assay and courtship behaviours of *Drosophila* (Sun and Heckscher, 2016; Nichols *et al.,* 2012). In this study, a simple larval locomotion assay was used to probe for lncRNA mutants impact this task (Nichols *et al.,* 2012). Although the results did not reflect any difference between the wildtype and the larvae from the viable, homozygous null mutant flies, we examined few escapee larvae of lncRNA-9 mutants and found them to display uncoordinated, paralytic movements. Further functional studies will be required to determine the cause of the locomotion defect and also the extension of lncRNA roles in the development of the *Drosophila* nervous system. Another possible reason for the locomotion defect could also arise due to a compromised cuticle and subsequent examination of the cuticle in the mutant will allow us to determine if the cuticle is fully developed.

From the phenotypic characterisation of the lncRNAs in this study, only two of the candidates displayed a homozygous lethal phenotype whereas the genomic deletions of the other lncRNA sequences failed to display any observable phenotype. This result is in line with other studies that have been performed using deletion mutagenesis of lncRNA candidates and a lack of noticeable phenotype was observed (Wei *et al.,* 2019; Goudarzi *et al.,* 2019;

Schor *et al.,* 2018). It is likely that many lncRNAs are not master regulators but instead function as buffering mechanism of gene expression (Wei *et al.,* 2019). In-depth studies of these candidates that do not display detectable phenotypes will allow us to test this hypothesis. Furthermore, during our investigation to examine the impact of the loss of the lncRNA candidates, the experiments that have been carried out were mainly based on overt features. However, it is possible that subtle defects or phenotypes were not detected. LncRNA-9 can be considered a bona fide lncRNA in *Drosophila* as the phenotype has been recapitulated using a second method, whereas further experiments have to be carried out to validate the phenotype observed for lncRNA-3. In addition, it is important to note that although computational algorithms were used to assess the protein-coding potential of these transcripts, experimental validation is still required to confirm that these lncRNAs are not translated into short polypeptides (Chew *et al.,* 2013; Yeasmin *et al.,* 2018).

Considerations from the loss-of-function outcomes

LncRNAs have emerged as an important class of regulators of gene expression during development and each of them have been known to possess unique features, characteristics and importance (Goff and Rinn, 2015; Kopp and Mendell; 2018). LncRNAs can be broadly divided into three groups based on their mechanisms of action from previous studies (Petruk *et al.,* 2006; Sauvageau *et al.,* 2013; Dimitrova *et al.,* 2014; Andergassen *et al.,* 2019). The first group include lncRNAs whose RNA product is important and contributes to gene regulation (e.g. *Fendrr*). The second group comprises of lncRNAs that do not generate a functional product. Instead, the initiation of transcription along the genomic loci is important for the expression of downstream genes. A good example is the act of transcription through the *Drosophila bxd* lncRNA that is needed to repress the expression of the downstream *Ubx* gene (Petruk *et al.,* 2006). Lastly, the third group refers to the underlying DNA elements within the lncRNA locus that are more likely to be the source of regulatory activity than the actual lncRNA product itself. Different perturbation techniques can be employed to study these lncRNA properties. Each of these methods have its pros and cons which will allow us to examine if the resulting phenotype is triggered by the loss of the RNA product or the underlying DNA elements that are embedded within the genomic locus of the lncRNA.

The first strategy that was used in this study made use of CRISPR/Cas9 to delete the lncRNA promoter so as to probe for the functional relevance of each lncRNA candidate in the list during *Drosophila* embryonic development. The TSS for each of lncRNA candidates were identified using RAMPAGE and the subsequent removal of the promoter region would

prevent the transcription of the lncRNA. I have taken a broad approach as the first step by removing the promoter region upstream from the TSS as well as part of the lncRNA sequence downstream from the TSS for each of the lncRNA candidates. This method was chosen as it would allow us to test if the lack of the lncRNA transcription and/or the lncRNA product itself would contribute to a phenotype. While most of the mutants that have been successfully generated did not exhibit any overt developmental defects, two of the lncRNA were found to be required for viability. I have also further validated the mutant phenotype using several methods to examine for possible off-targeting effects by CRISPR.

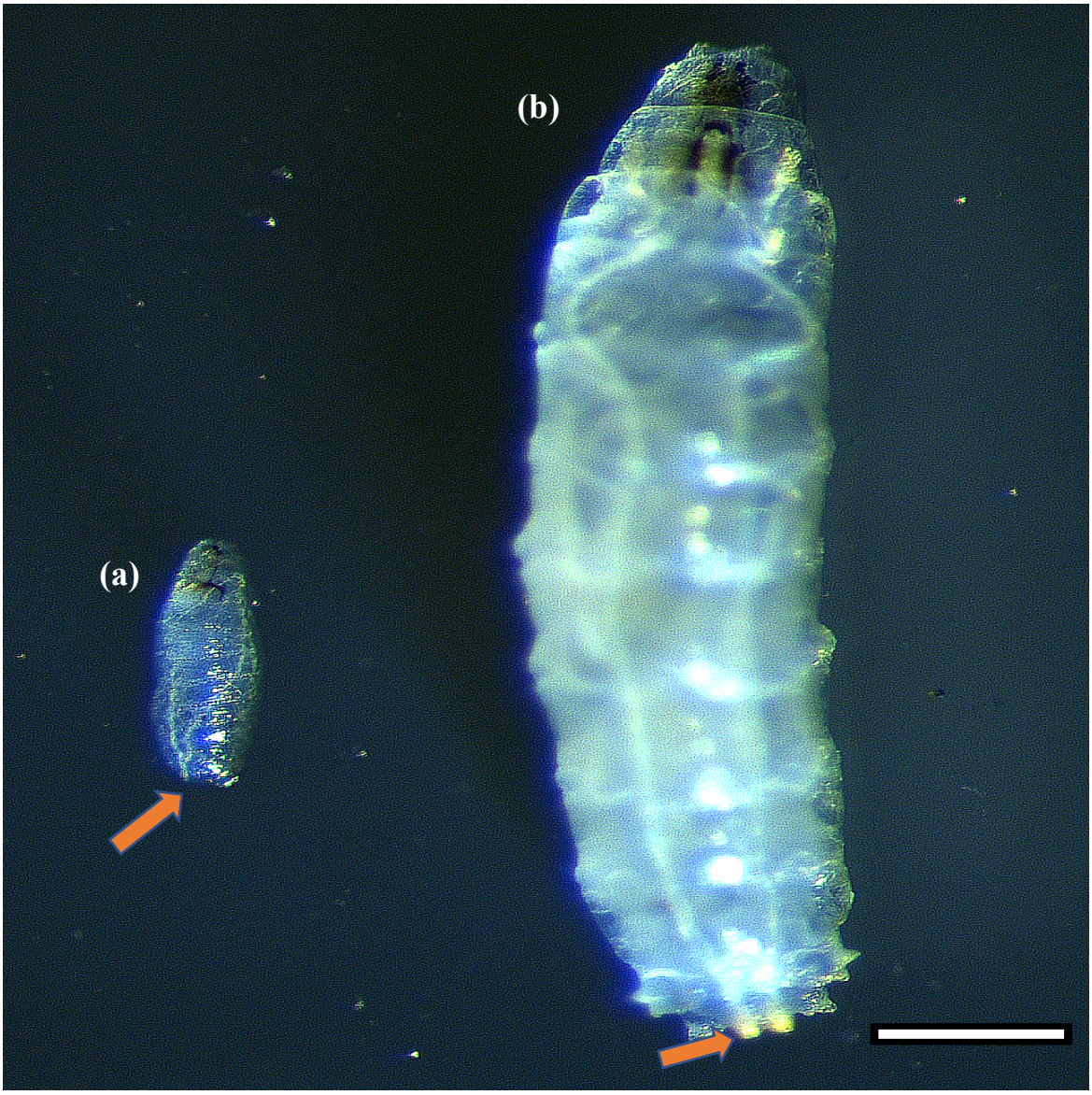
Nevertheless, while I have addressed the off-target effects of the perturbation method used, additional studies will be needed to understand if the phenotype has risen due to the lack of transcription or the lncRNA product itself. To answer this question, the integration of a transcriptional stop signal at the beginning of the lncRNA transcript using CRISPR/Cas9 was adopted. This strategy was chosen as it generates a LOF mutation through the insertion of a SV40 polyadenylation signal (polyA). Due to the premature cleavage and subsequent polyadenylation, the production of a full lncRNA transcript was inhibited. As the promoter region would remain intact using this method, any phenotype observed would be due to the lncRNA transcript itself. Indeed, I have observed a similar phenotype for the lncRNA-9 mutant which implies that the lncRNA product itself is important for its function. On the other hand, due to the unsuccessful attempts to generate founders for lncRNA-3 using this method, it was impossible to tease apart the two elements.

While the two methods I used allowed me to identify the functional lncRNA product, this study did not examine the effect of possible underlying DNA elements. The presence of gene regulations by enhancers or insulators within the lncRNA loci have been reported to be common. The removal of these components while attempting to mutate lncRNAs have also been known to affect the expression of neighbouring genes (Engreitz *et al.,* 2016; Dimitrova *et al.,* 2014; Groff *et al.,* 2016). Nevertheless, further validation using methods such as RNAi or CRISPRi that do not affect the genomic locus will be needed to completely exclude that the observed effects are due to the deletion of undetected DNA elements present in the lncRNA locus. Subsequently, deletions could be made along the lncRNA transcript so as to identify important regions containing these DNA elements.

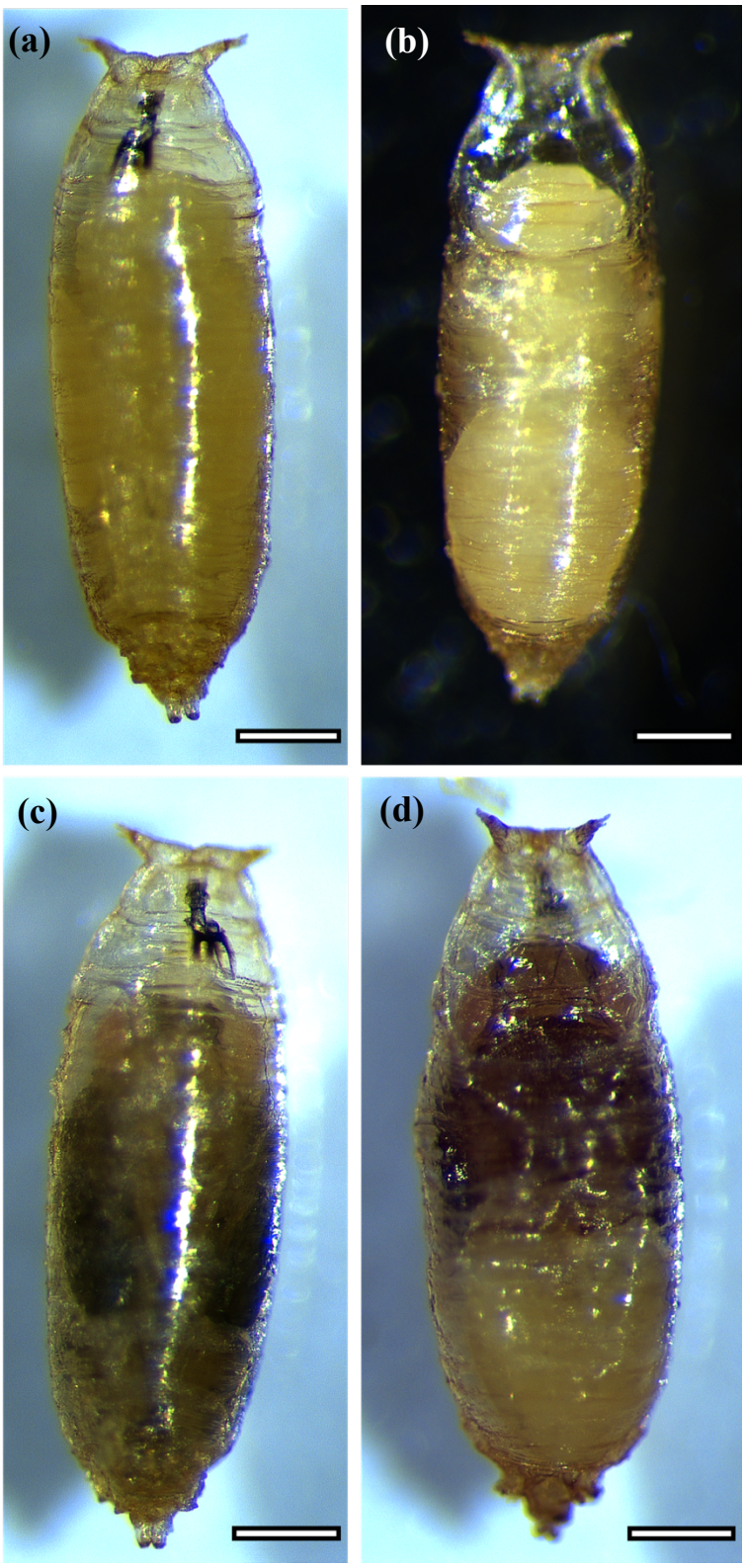
Conclusion

Considered as a whole, we have begun to appreciate the functions of lncRNAs using model organisms such as *Drosophila melanogaster* and this could open up many doors due to their specific spatio-temporal expression and undetermined functions. In this study, we made use of RAMPAGE data and identified a total of 3,973 lncRNA TSCs in *Drosophila melanogaster*. A total of 631 of these loci have displayed high sequence conservation and expression profiles across the five *Drosophila* species. This evidence has hinted that these lncRNA loci carry important biological functions and are ideal candidates for further experimental scrutiny into their contributions to developmental processes. Using genome editing methodologies, I have examined the functional relevance of thirteen of these lncRNA candidates and proven that two of them yield in vivo phenotypes when deleted. I have showed that the function of lncRNA-9 is likely to be conveyed by the RNA transcript. From the outcomes of the vivo assays, due considerations have to be made during the experimental design so as to decipher the fitness of each lncRNA loci as the bulk of the lncRNA candidates do not display any gross phenotypes. Although these mutants generated did not display any developmental defects nor were found to be essential for viability, the results do necessarily imply that they are not functional. Due to the limitations of our assays, further studies are required to determine their functional roles. Although it would be easy to classify them as functionally redundant like the *roX* RNAs, recent findings have shown that it is not entirely true for the *roX* RNAs as *roX1* and *roX2* lncRNA were found highly expressed in different developmental stages, suggesting that may have different role during dosage compensation (Kim *et al.,* 2018). As the lncRNAs selected were found to have a conserved expression profile across five *Drosophila* species, it would be interesting to find out if mutagenesis of these transcripts in a different genetic strain would produce the same results. This would allow us to understand if lncRNAs were strain-specific as most of the functional studies were only carried out in *Drosophila melanogaster*. While I was unable to generate any lncRNA mutants for eight of the lncRNA candidates in the list during my study, it has now been made possible with the advancement of the CRISPR/Cas9 technology which include mutagenesis at tissue-specific levels (Port *et al.,* 2020). Lastly, multiple lines of evidences are needed when investigating each lncRNA loci as the phenotype observed could be caused by a loss of the RNA transcript or the underlying DNA elements in the genomic loci.

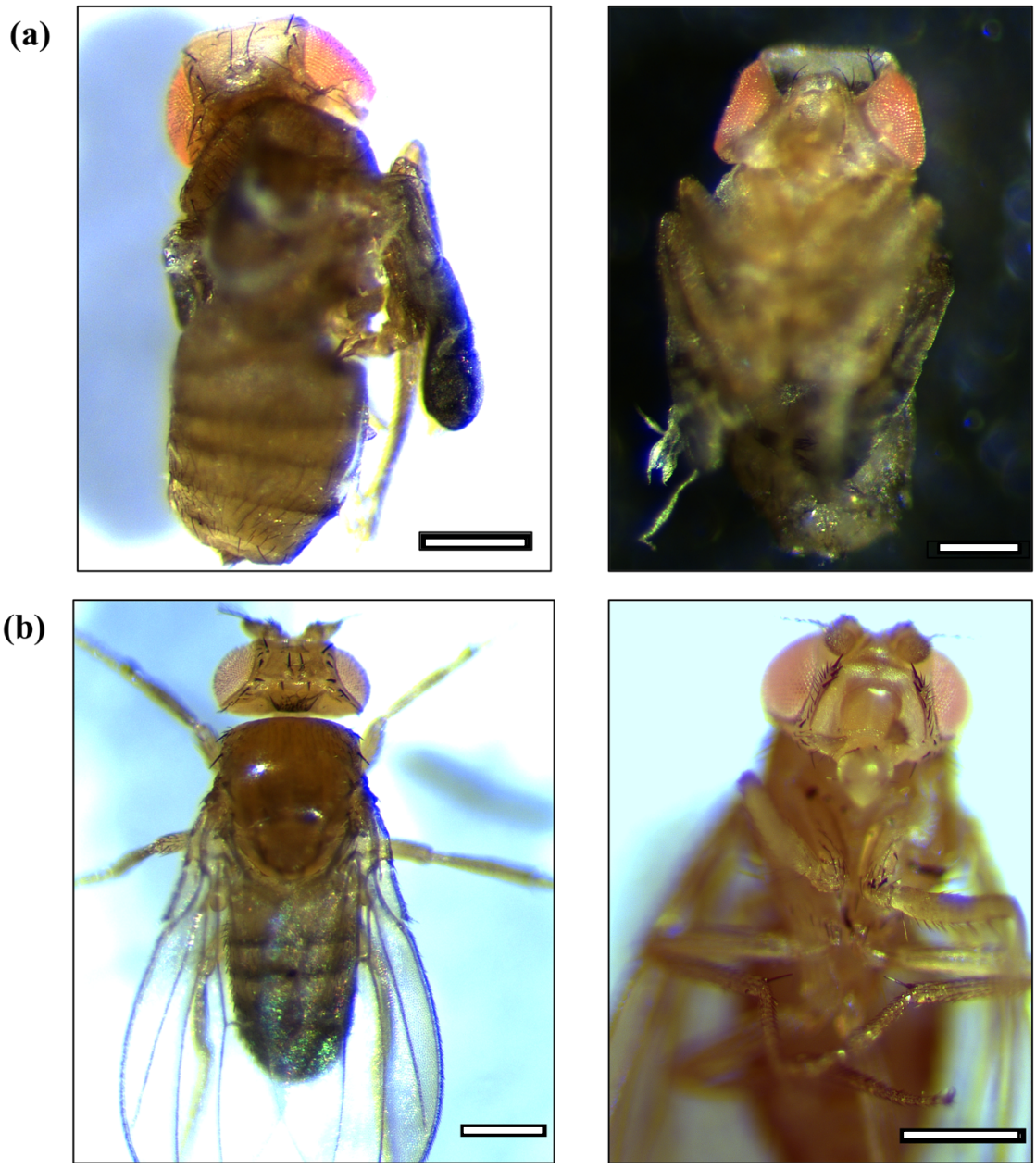
Supplementary Figures



Supplementary Figure 1: Size difference between lncRNA-9 escapees and wildtype larvae. *Drosophila* larvae have three instar stages. The first and second instar lasts one day each whereas the third instar stage persists for two more days. A comparison of the (a) lncRNA-9 escapee with a wildtype larvae (right) 36 hr after hatching from the embryo is shown. The larvae have been arranged with the anterior end of the larvae at the top and the posterior end at the bottom of the image for both larvae. The trachea and mouth hook can be clearly seen in both larvae. One of the most striking features is the absence of posterior spiracles (orange arrow) and a much smaller size in the lncRNA escapee. In addition, the wildtype larvae entered the second instar stage by this time and has a mouth hook that appears to be much more developed compared to the mutant in the first instar stage. The escapee arrest at L1 stage and no movement was observed even though a brush was used to gently push them. The majority of the wildtype larvae were in the second instar stage at this point. (Scale bar, 1 mm.)


(a)
(b)
(c)
(d)

Supplementary Figure 2: Difference in pupal development stages found in lncRNA-3 homozygous null mutants.
The images illustrate the pupal development of lncRNA-3 homozygous mutants during the stages of pupariation to adult eclosion. LncRNA-3 mutants were manually separated from their heterozygous siblings during embryonic stage. The embryos were allowed to develop at 25°C and the images were taken 16 days after egg laying. All of the wildtype and heterozygous siblings had emerged into adults by this time. Compared to wildtype and heterozygous siblings, lncRNA-3 homozygous null mutants had vastly different rate of development even though the embryos were collected within a 4 hr window. (a,c) Ventral view of the lncRNA mutant. (b,d) Dorsal view of the mutant. (a,b) The mutant is in the early stages of pupation and have detached from the case. The body have turned yellow and the head has everted. (c,d) The mutant has completed most of the stages of pupal development with the formation of eyes and wings. A total of 100 homozygous mutant pupae were examined. Although all of them were incubated for a total of 30 days, none of the homozygous mutants completed eclosion and emerged from pupation. (Scale bar, 0.5 mm.)



Supplementary Figure 3: Difference in the anatomical morphology observed in lncRNA-3 homozygous null mutants.
Pupal lethal phenotypes associated with lncRNA-3 mutations. Dorsal and ventral view of (a) Dissected lncRNA-3 homozygous mutants in late pharate adult stage and (b) wildtype adult. The majority of the homozygous null mutants were arrested at the pharate adult stage at the end of 30 days. A difference in the development of the head and appendages were observed when the mutants were compared to a wildtype adult. The mutants were found to be unable to extend head fully at the end of development. (Scale bar, 0.5 mm)

For supplementary figures 4 – 16:

Anatomical images of homozygous lncRNA null mutants except for supplementary figure 5 (lncRNA-3***)*** and supplementary figure 8 (lncRNA-9) whereby only heterozygous mutants were viable. Using wildtype flies shown in supplementary figure 4 as reference, we compared 6 different anatomical features and we did not observe any difference between the wildtype adult flies and lncRNA mutants. All scale bars represent 500 µM. The images were arranged in the following order:

(a) Examination of the side view of mutant fly with arrows pointing to the (1) head, (2) legs and (3) wings. For the head, we examined the overall shape of the head and eyes as well as the antenna. For the leg, we checked that there were 6 legs and the fly was able to walk properly. For the wings, they had to be developed fully and extend beyond the body of the fruit fly.

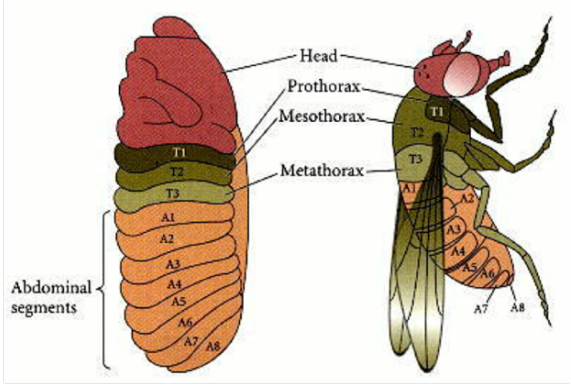
(b) Top-down view of the mutant fly with arrow pointing to the (4) thorax, (5) pair of halters and (6) abdominal segments. For the thorax region, there should be the prothorax, mesothorax and metathorax. The flies would have a pair of haltere instead of a wing-like structure seen in *Bxd* mutants. We counted the abdominal segments to ensure that the flies have 8 abdominal segments.

(c) CRISPR/Cas9 induced HDR with successful integration of the RFP-carrying donor vector. Compared to wild type flies in supplementary figure 4, a strong RFP signal could be seen under the fluorescence microscope in the eyes of flies with successful genomic integration of the donor plasmid. In the wildtype fly, the RFP signal should be absent.

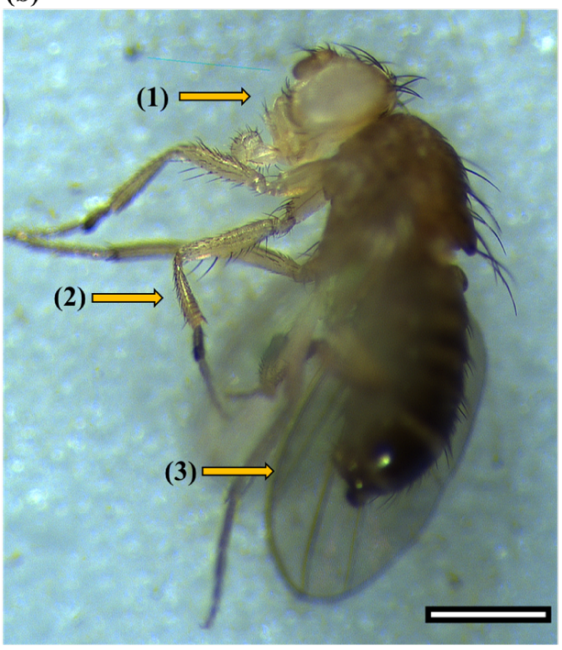
(d) Expression profiles throughout embryogenesis of 5 *Drosophila* species. LncRNA expression was normalized across the different *Drosophila* species and plotted using an arbitrary developmental time course. These expression profiles were generated by Dr. Philippe Batut.

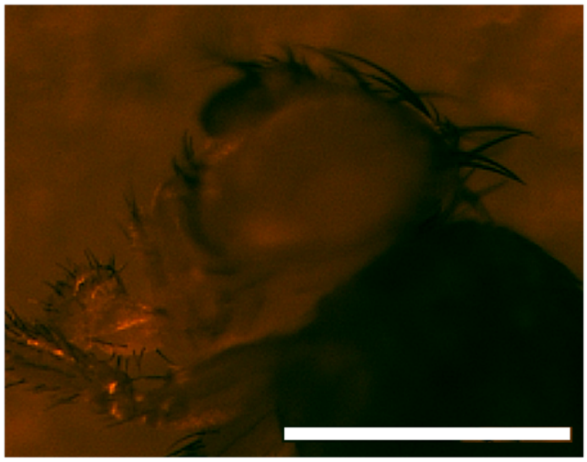
(e) Genomic loci of each lncRNA. The orange bar indicates the deleted region and replacement with a 3xP3-RFP cassette.

(a)

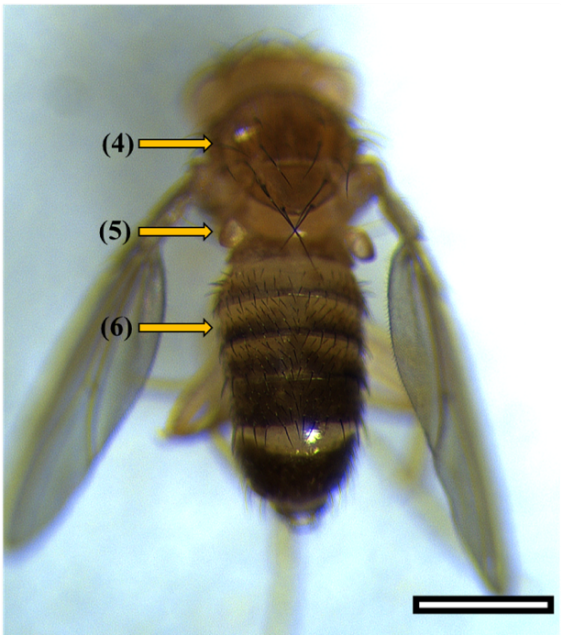

Head
Prothorax
Mesothorax
Metathorax
T1
T2
T3
A1
A2
A3
A4
A5
A6
A7
A8
Abdominal segments

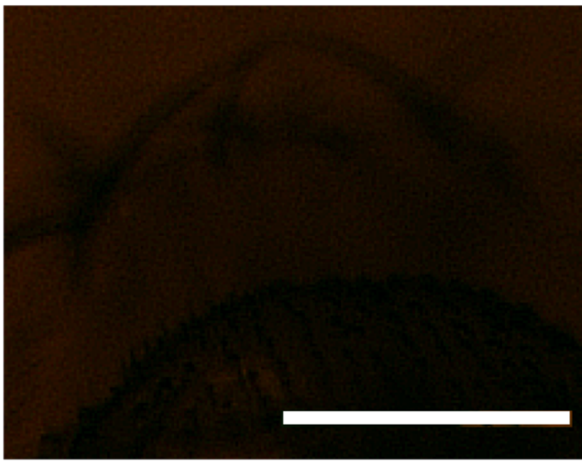
(b)


(1)
(2)
(3)

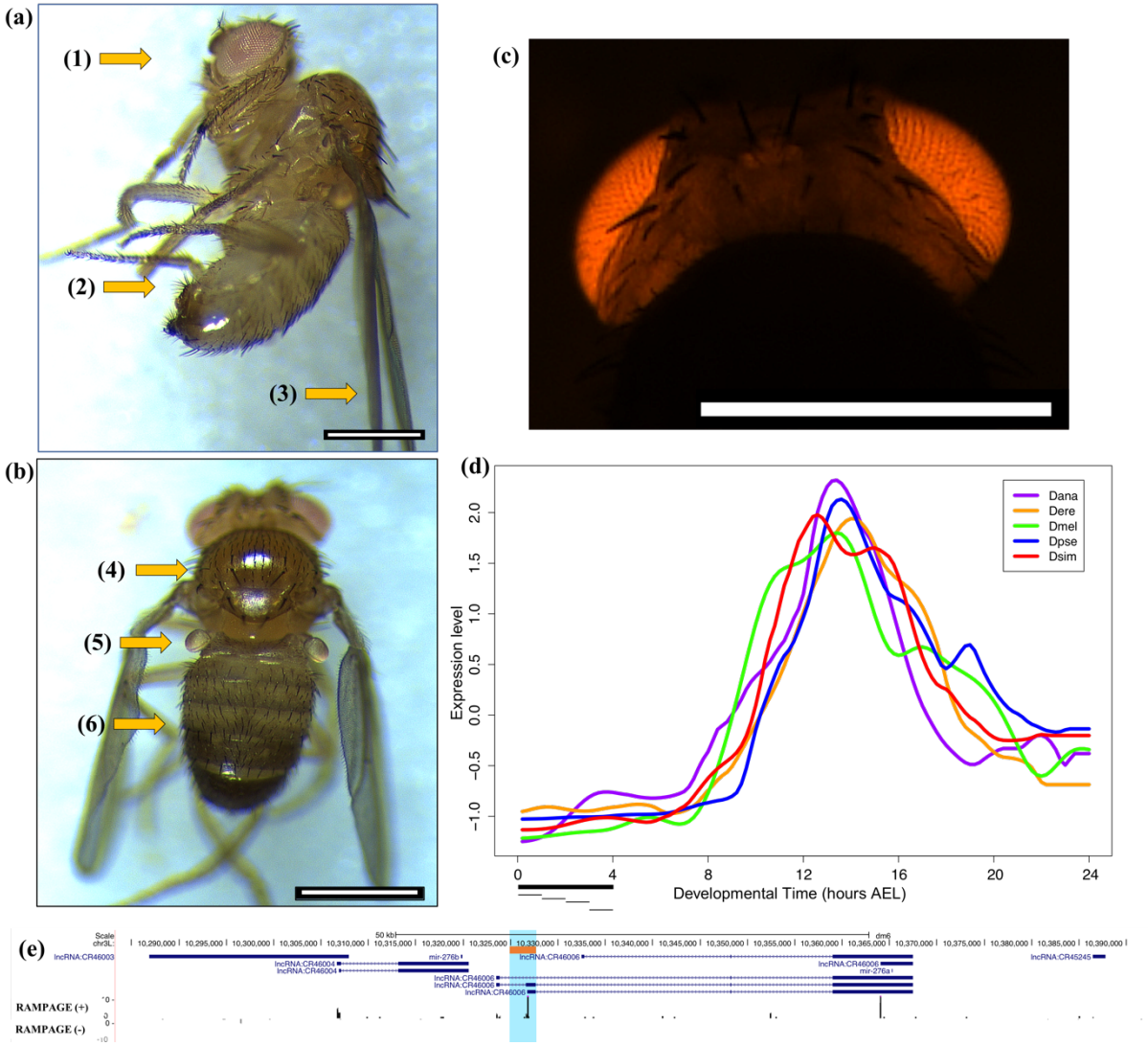


(c)

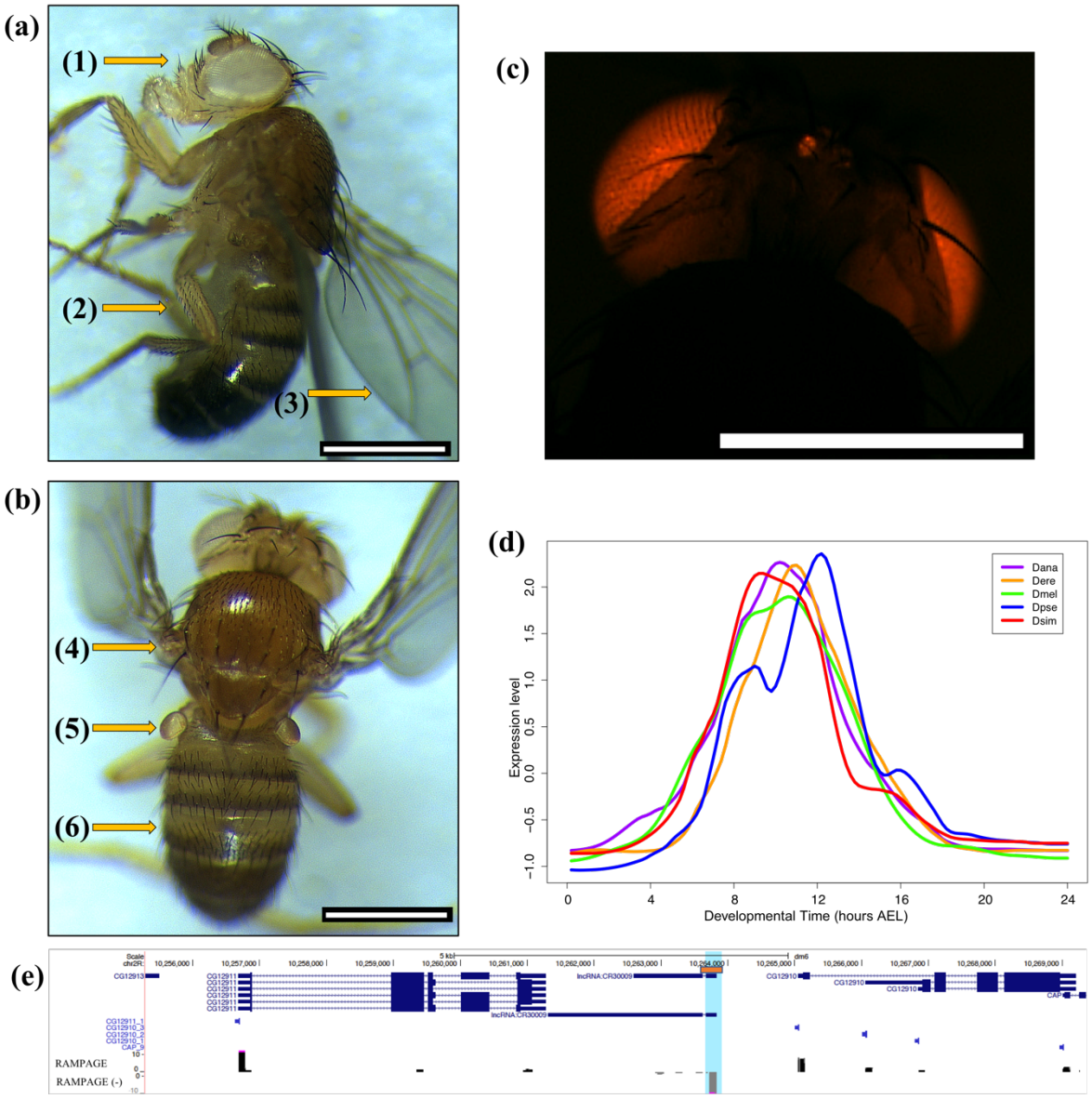

(4)
(5)
(6)



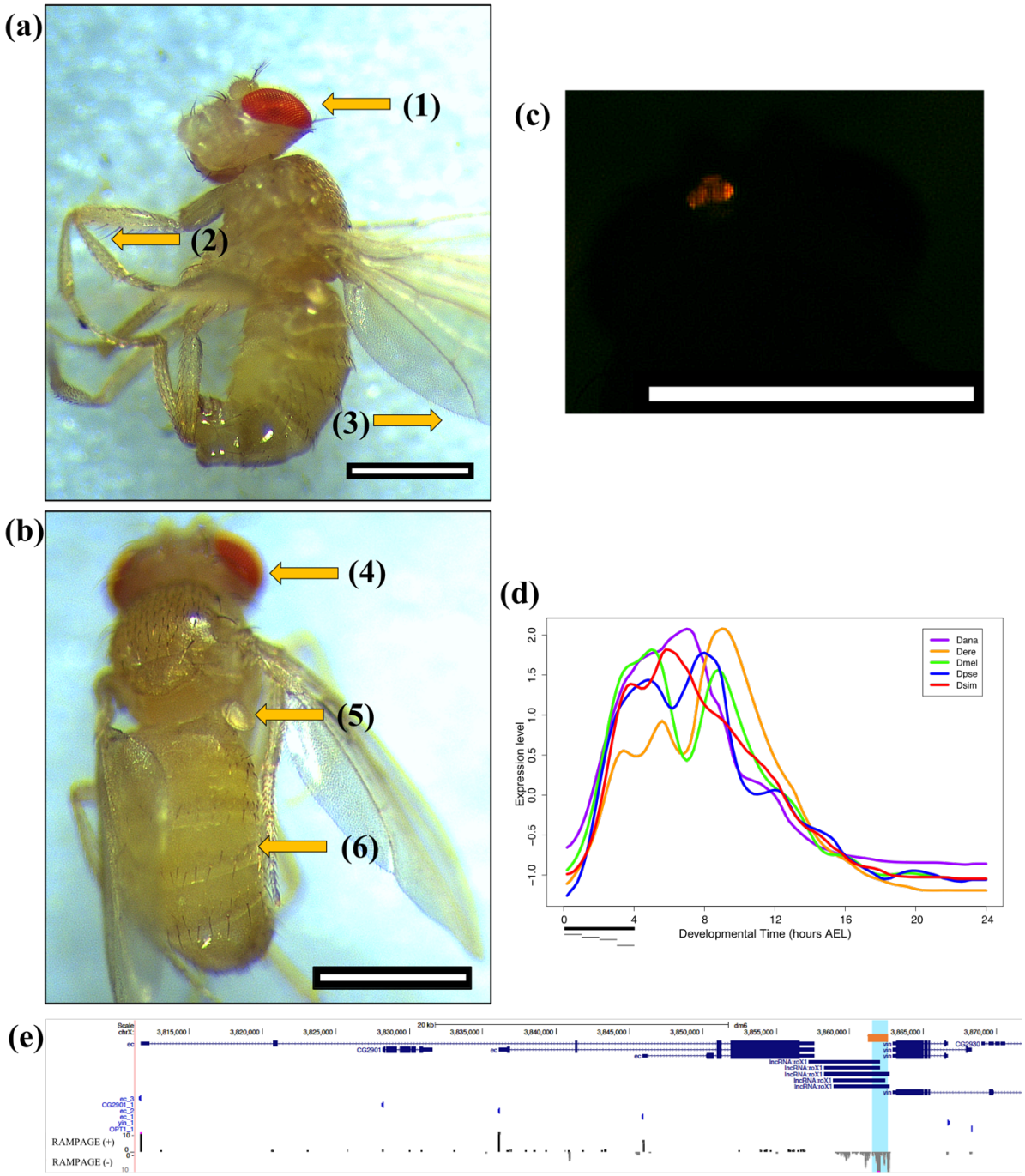
Supplementary Figure 4: Control w^{1118}
(a) Schematic of the *Drosophila* adult body axis as reference.
An adult fly consists of a head region, a thoracic segment (T1-T3) and an abdominal segment (A1-A8). (b) Lateral view of a control w^{1118} adult male. Examination of the lateral view of mutant fly with arrows pointing to the [1] head, [2] legs and [3] wings. For the head, we examined the overall development of the shape of the head, mouth, eyes as well as the antenna. For the leg, we checked that there were 6 legs and the fly was able to walk properly. For the wings, they had to be developed fully and extend beyond their body of the fruit fly. (c) Dorsal view of the control w^{1118} adult male.
[4] The notum that is composed of the anterior scutum and posterior scutellum. [5] The haltere is made up of 3 components. They are the scabellum, the pedicel and the capitellum. [6] A normal adult fruit fly would have 8 abdominal segments (A1-A8). (Scale bar, 0.5 mm)


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)

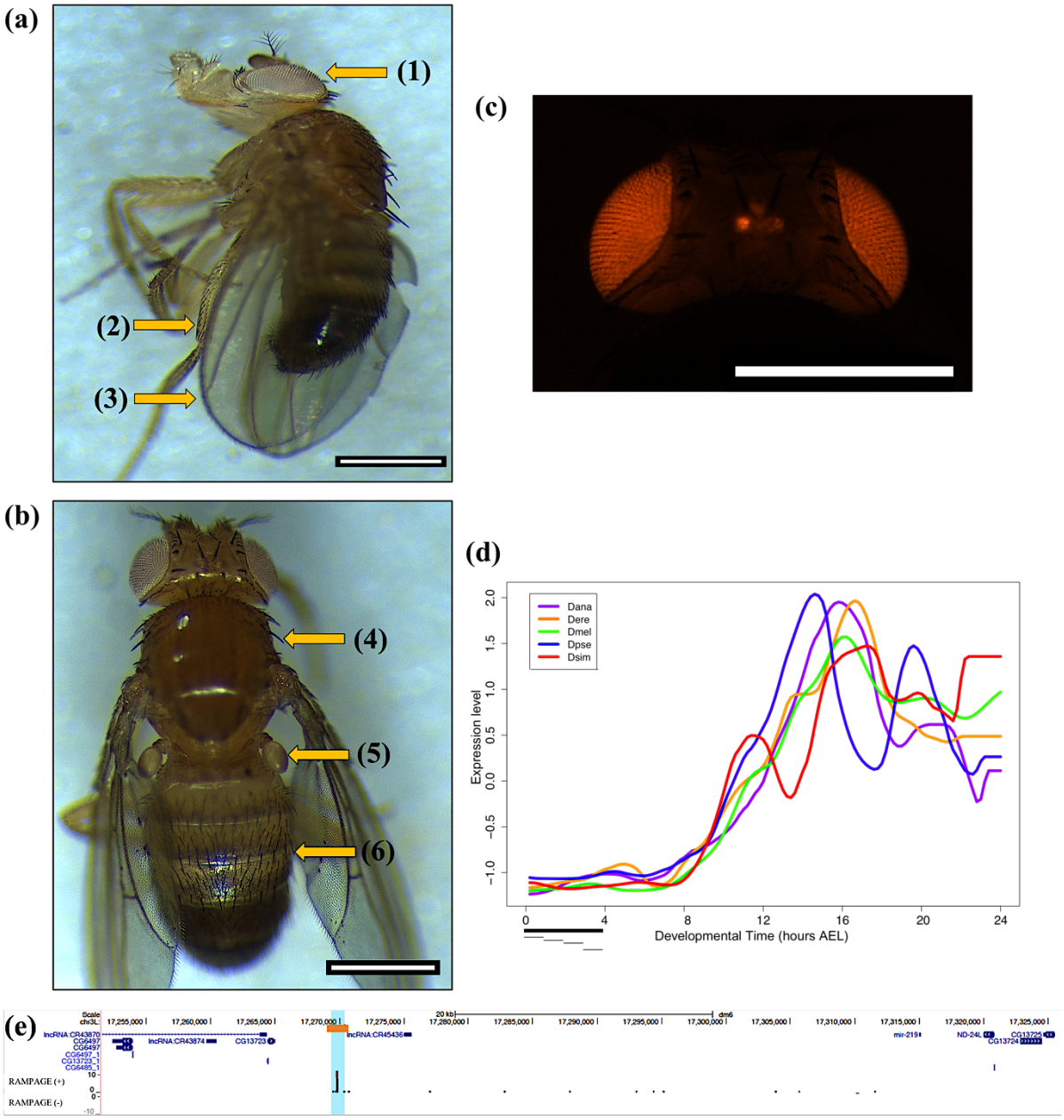
Supplementary Figure 5: lncRNA-3 (*FBgn0265415_02*) heterozygous mutant
Homozygous null mutants for this lncRNA were pupal lethal and no homozygous adults were collected. LncRNA-3 heterozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 heterozygous males and 10 heterozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-3 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-3, the highest expression level of the lncRNA in *Drosophila melanogaster* is at 14 hr after egg laying. (e) Genomic locus of lncRNA-3 as visualised on the UCSC genome browser. Transcripts that have been annotated on Flybase are shown with dark blue bars with m*icroRNA-276b* located upstream of the lncRNA and *lncRNA:CR45245* downstream of the transcript. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows the promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-3 is expressed from the plus strand of the DNA and has been annotated on Flybase as *lncRNA:CR46006* with 5 different isoforms. The new annotation of these transcripts indicated that the lncRNA overlaps with *microRNA-276a*. A pair of gRNAs was used to remove the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr3L: 10,325,557 – 10,327,321.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE
RAMPAGE (-)

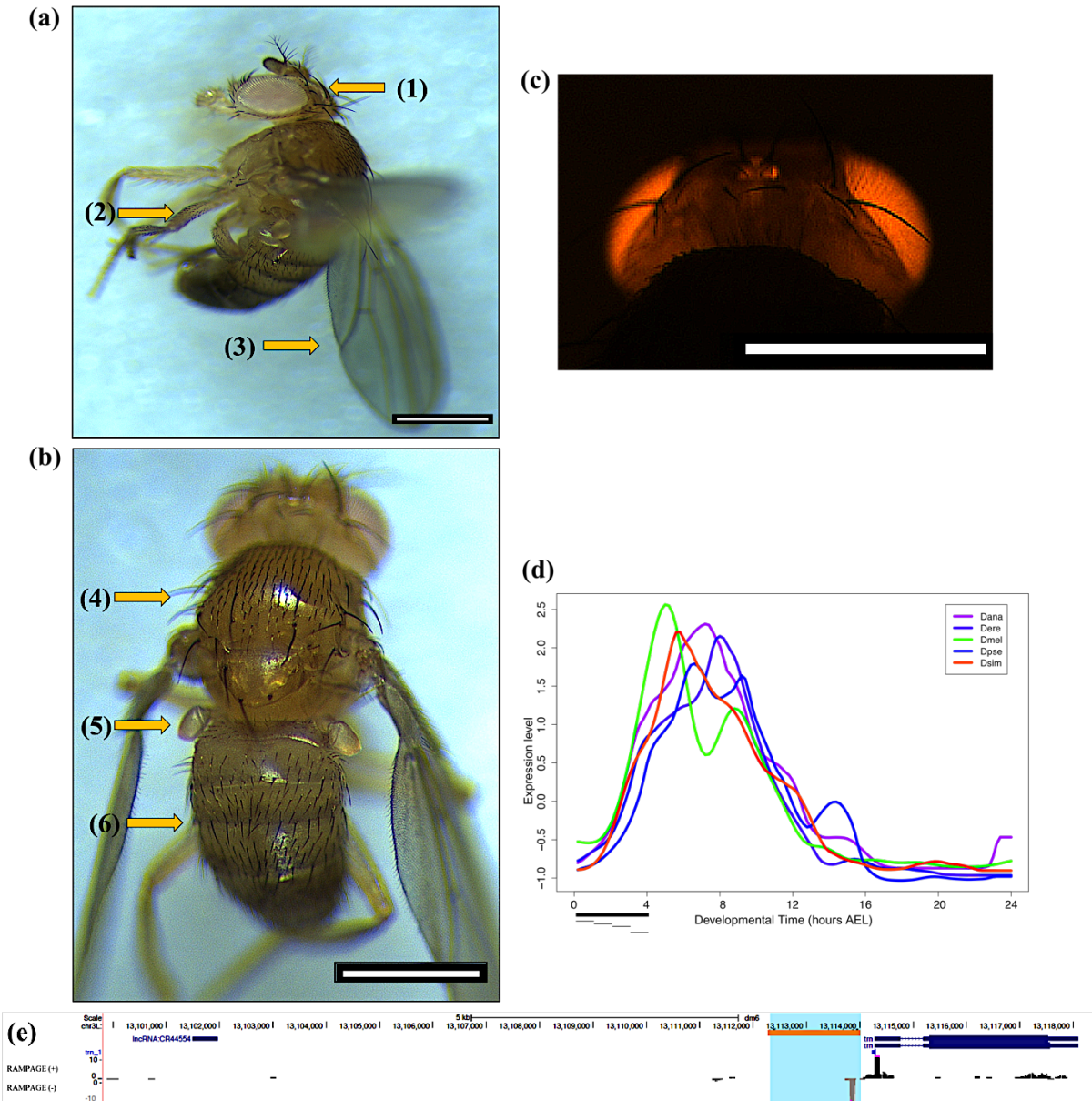
Supplementary Figure 6: lncRNA-4 (*FBgn0050009_02*) homozygous mutant.
LncRNA-4 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image show eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic loci. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-4 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-4, the highest expression level of the lncRNA in *Drosophila melanogaster* is at 10 hr after egg laying. (e) Genomic loci of lncRNA-4 as visualised on the UCSC genome browser. Transcripts that have been annotated on Flybase are shown with dark blue bars with *CG12911* and *CG12910* as the two closest genes to the lncRNA. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-4 is expressed from the negative strand of the DNA and has been annotated on Flybase as *lncRNA:CR30009* with two different isoforms. A pair of gRNAs was used to excise the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr2R: 10,263,671– 10,263,910.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)

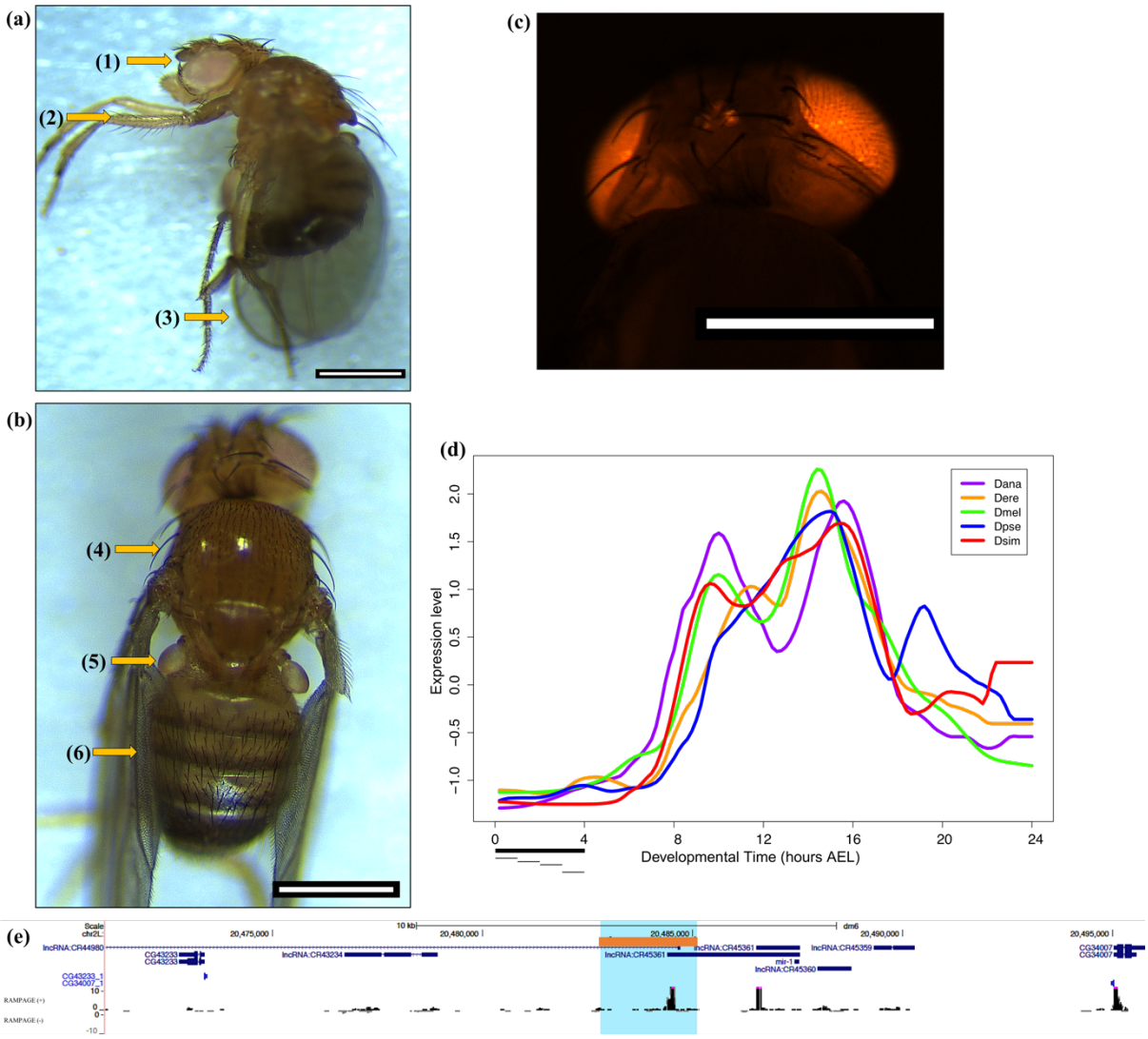
Supplementary Figure 7: lncRNA-5 (*FBgn0019661_05, lncRNA: rox1*) homozygous mutant. *LncRNA: rox1* has been published in the literature and it is one of the most studied lncRNA in *Drosophila*. The lncRNA has been reported to be redundant with *lncRNA: rox2* (Quinn *et al.,* 2016) and we have used this transcript as a positive control. Homozygous *rox1* mutants were viable and this result has been reported in previous studies (Bassett *et al.,* 2014; Quinn *et al.,* 2016). (a,b) Homozygous lncRNA mutant with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image show eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus. Due to the red pigmentation of the adult eye, the RFP signal can only be visualised from the *Drosophila* Ocelli. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-5 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-5, the highest expression level of the lncRNA transcript in *Drosophila melanogaster* is detected at the early stages of embryonic development. Two expression peaks have been identified from the expression data (5 hr, 9 hr after egg laying). (e) Genomic loci of lncRNA-5 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars. *Echinus* (*ec*) and *yin* are two protein-coding genes that are the closest to the lncRNA transcript. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-5 is expressed from the negative strand of the DNA and has been annotated on Flybase with five different isoforms. A pair of gRNAs was used to delete the promoter region of the lncRNA as shown (orange bar, blue highlight) and HDR was used to insert a 3xP3-RFP cassette into the locus. The region excised is from chrX: 3,861,568 – 3,862,573.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)

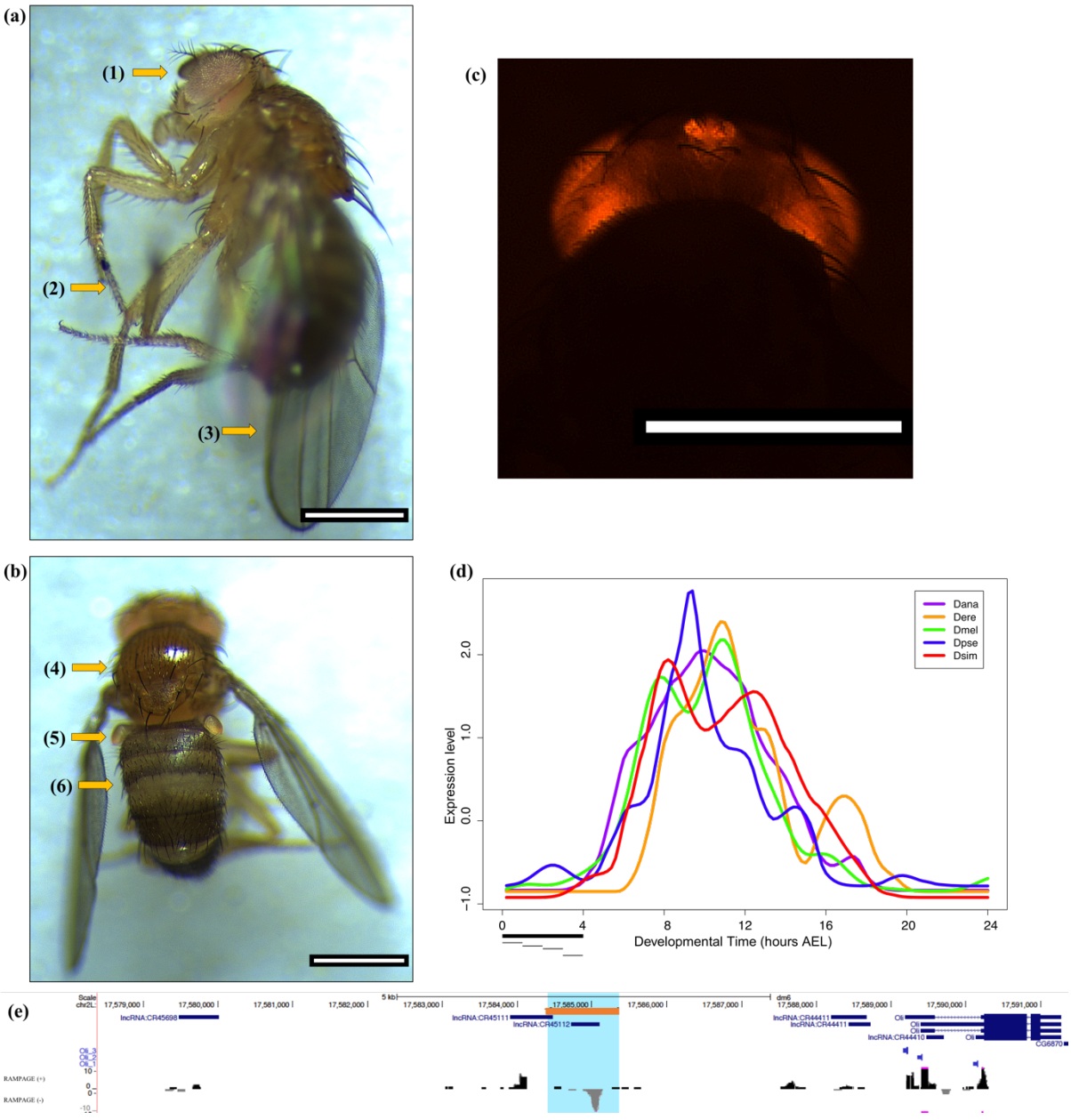
Supplementary Figure 8: lncRNA-9 (*RAMPAGEpeak_002227*) heterozygous mutant.
The majority of the homozygous lncRNA null mutants were embryonic lethal. LncRNA-9 heterozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 heterozygous males and 10 heterozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic loci. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-9 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-9, the highest expression level of the lncRNA in *Drosophila melanogaster* is at 14 - 16 hr after egg laying. (e) Genomic locus of lncRNA-9 as visualised on the UCSC genome browser. Transcripts that have been annotated on Flybase were shown with dark blue bars with *CG13723* located upstream of the lncRNA whereas *lncRNA:CR45436* is downstream. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-9 is expressed from the plus strand of the DNA and it is unannotated on Flybase or the Ensembl genome database. A pair of gRNAs was used to remove the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr3L: 17,269,326 – 17,270,407.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
lncRNA:CR44554
RAMPAGE (+)
RAMPAGE (-)

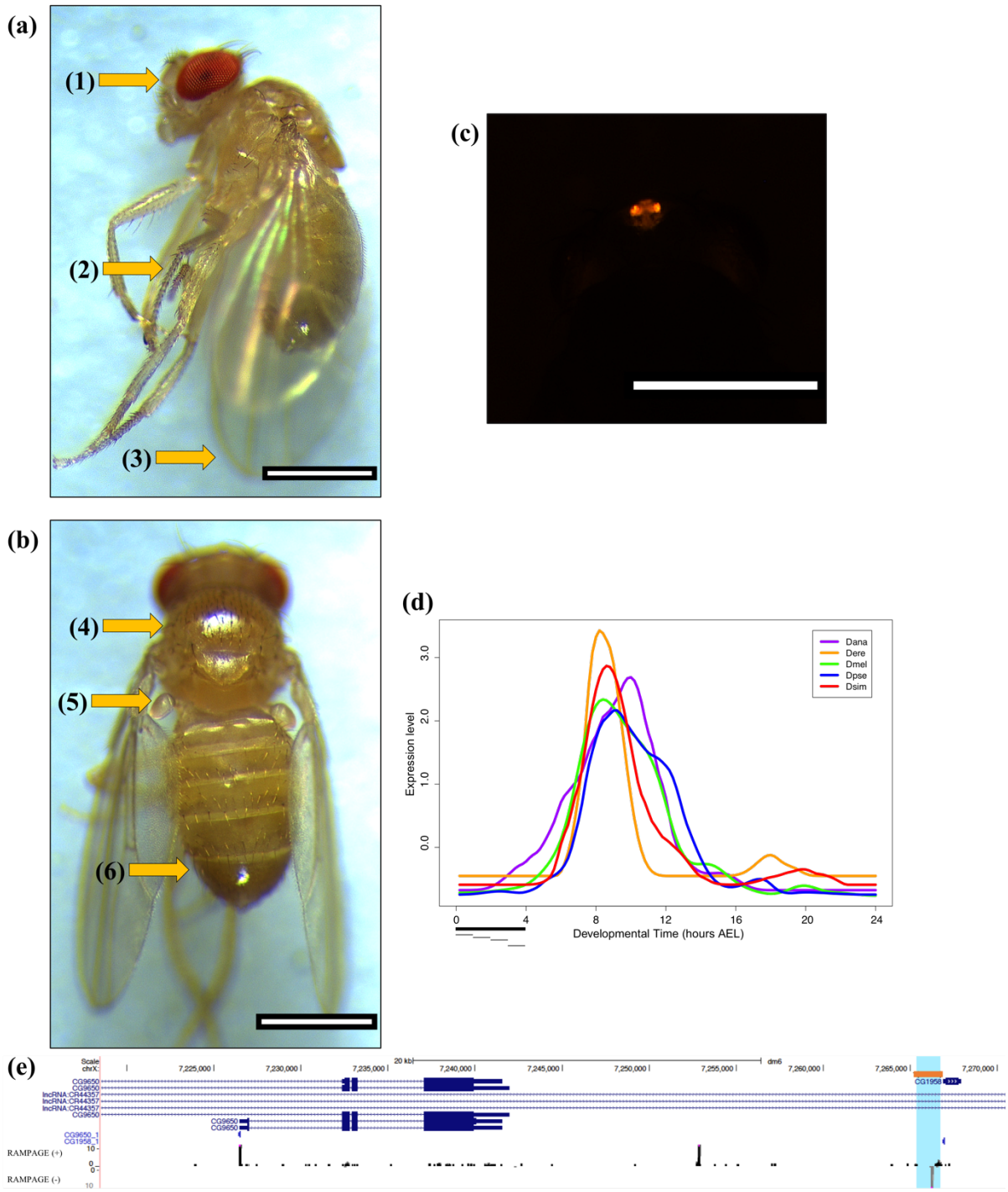
Supplementary Figure 9: lncRNA-10 (*RAMPAGEpeak_005280*) homozygous mutant.
LncRNA-10 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows the eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-10 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-10, the highest expression level of the lncRNA in *Drosophila melanogaster* is at 5 - 6 hr after egg laying (AEL) and a smaller peak was seen at a later time point (9 – 10 hr). (e) Genomic locus of lncRNA-10 as visualised on the UCSC genome browser. Transcripts that annotated on Flybase are shown with dark blue bars with *lncRNA:CR44554* and *tartan* (trn) as the two closest genes to the lncRNA. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-10 is expressed from the negative strand of the DNA and has not been annotated on Flybase or the Ensembl genome assembly. A pair of gRNAs was used to excise the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr3L: 13,112,379 – 13,113,965.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)

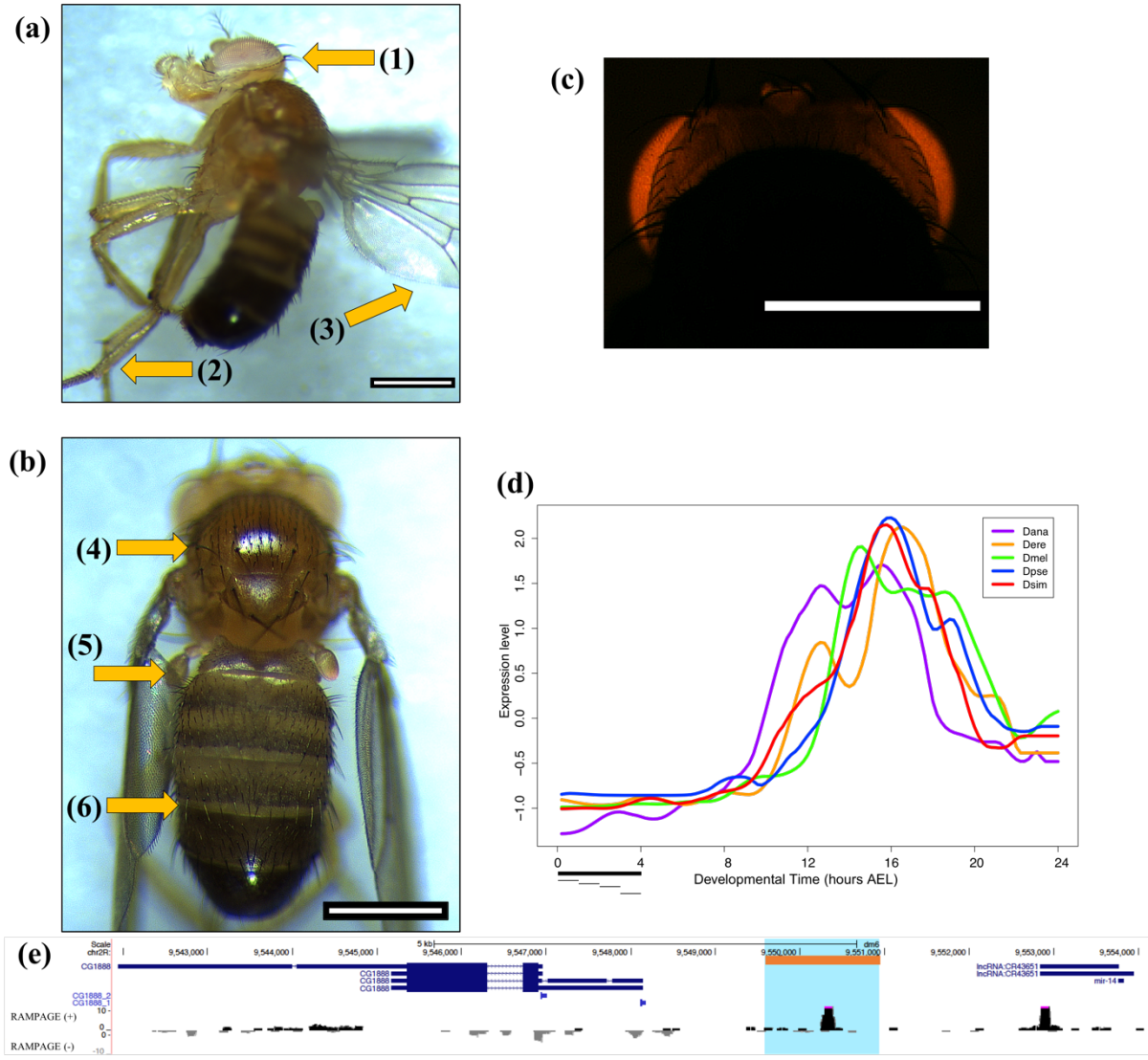
Supplementary Figure 10: lncRNA-11 (*RAMPAGEpeak_006440*) homozygous mutant. LncRNA-11 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic loci. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-11 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-11, the highest expression level of the lncRNA in *Drosophila melanogaster* occurred around 14 - 15 hr although a smaller peak of expression was detected earlier at around 10 hr after egg laying (AEL). (e) Genomic loci of lncRNA-11 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars. *LncRNA:CR43234* is located upstream of the lncRNA transcript whereas *lncRNA:CR45360* is located immediately downstream of lncRNA-11. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-11 is expressed from the plus strand of the DNA and is annotated on Flybase as *lncRNA: CR45361* with 2 different isoforms. A pair of gRNAs was used to excise the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr2L: 20,483,035 – 20,485,071.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
2.0
1.0
0.0
−1.0
0
4
8
12
16
20
24
Developmental Time (hours AEL)
(e)
Scale
chr2L:
17,579,000
17,580,000
17,581,000
17,582,000
5 kb
17,583,000
17,584,000
17,585,000
17,586,000
17,587,000
dm6
17,588,000
17,589,000
17,590,000
17,591,000
lncRNA:CR45698
lncRNA:CR45111
lncRNA:CR45112
lncRNA:CR44411
lncRNA:CR44410
Oli
CG6870
RAMPAGE (+)
RAMPAGE (-)

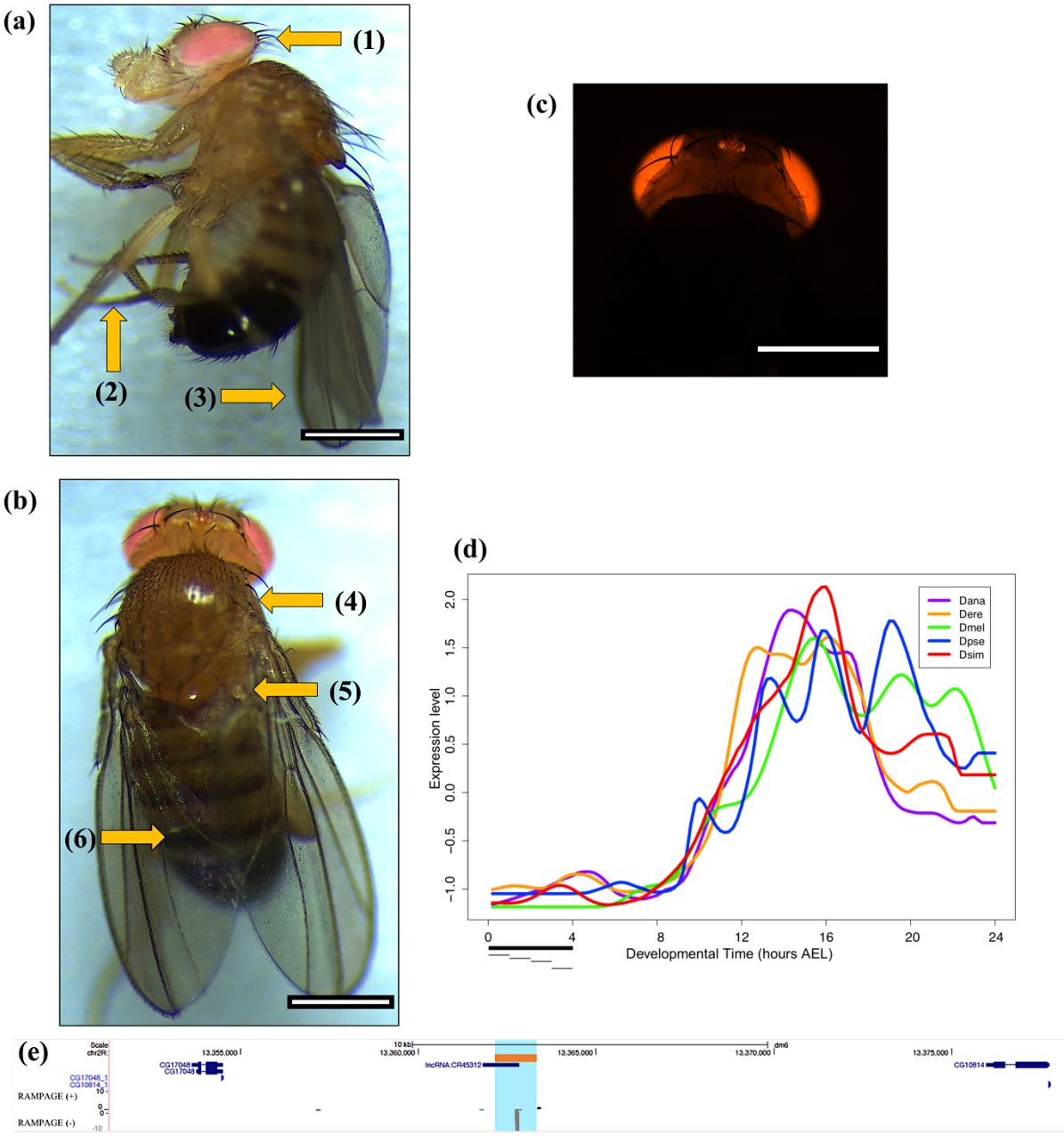
Supplementary Figure 11: lncRNA-12 (*RAMPAGEpeak_008667*) homozygous mutant.
LncRNA-12 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows eye phenotype of flies with successful integration of *3xP3-RFP* cassette at the lncRNA genomic loci. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-9 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-12, there were 2 peaks observed for the expression of the lncRNA transcript. The first peak occurred at 8 hr after egg laying and the second peak emerged at around 11-12 hr. (e) Genomic locus of lncRNA-12 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars with *lncRNA:CR44411* located upstream of the lncRNA whereas *lncRNA:CR45111* is located downstream. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-12 is expressed from the negative strand of the DNA and it is annotated on Flybase with one isoform. A pair of gRNAs was used to remove the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr2L: 17,584,436 – 17,585,359.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
0.0
1.0
2.0
3.0
0
4
8
12
16
20
24
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)

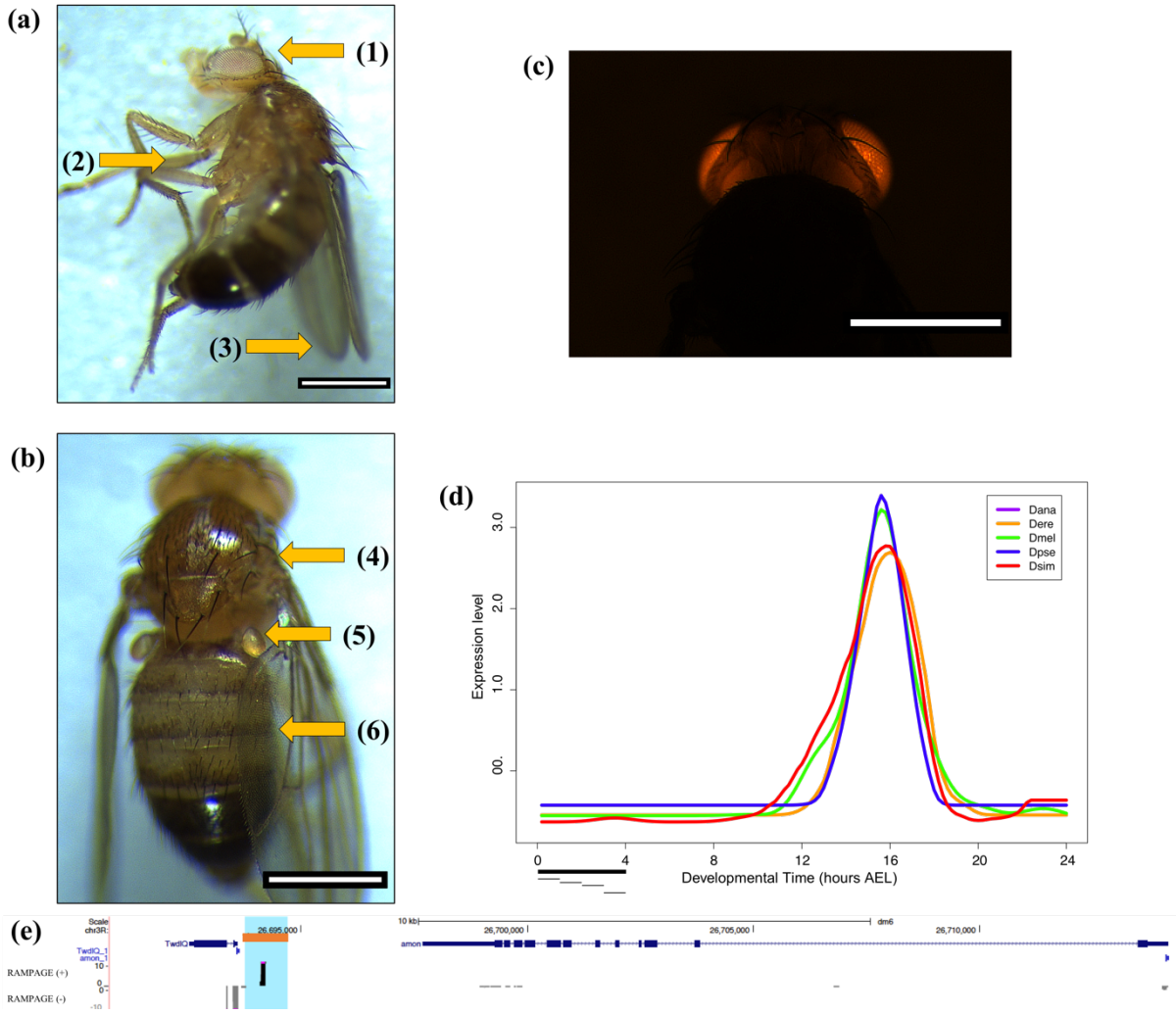
Supplementary Figure 12: lncRNA-13 (*RAMPAGEpeak_011922*) homozygous mutant. Homozygous null mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus. Due to the red pigmentation of the adult eye, the RFP signal can only be visualised from the *Drosophila* Ocelli. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-13 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-13, the highest expression level of the lncRNA transcript in *Drosophila melanogaster* was detected at 9 hr after egg laying. (e) Genomic loci of lncRNA-13 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars. *CG1958* and *CG9650* are two protein-coding genes that are the closest to the lncRNA transcript. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-13 is expressed from the negative strand of the DNA and the transcript has not been annotated on Flybase or the Ensembl genome assembly. The transcript is found to be located in the intronic region of another lncRNA transcript (*lncRNA: CR44357*). A pair of gRNAs was used to replace the promoter region of the lncRNA with a 3xP3-RFP cassette as shown (orange bar, blue highlight). The region excised is from chrX: 7,265,450 – 7,266,726.


(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)

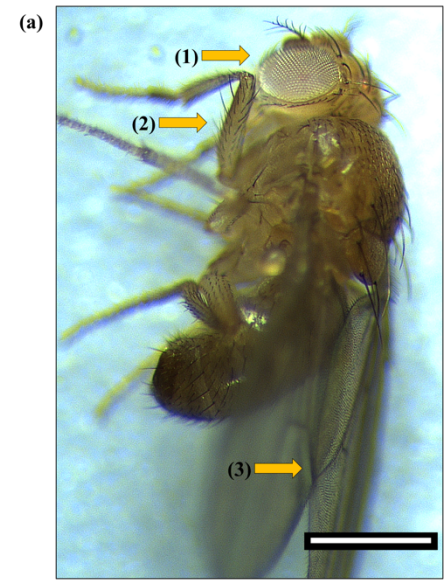
Supplementary Figure 13: lncRNA-14 (*RAMPAGEpeak_015813*) homozygous mutant.
A total of 10 homozygous males and 10 homozygous females were examined. LncRNA-14 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows the eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-14 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-14, the highest expression level of the lncRNA in *Drosophila melanogaster* is at 13 - 14 hr after egg laying (AEL). (e) Genomic locus of lncRNA-14 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars with *lncRNA:CR43651* and *CG1888* as the neighbouring genes to the lncRNA. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-14 is expressed from the plus strand of the DNA and has not been annotated on Flybase or the Ensembl genome assembly. A pair of gRNAs was used to excise the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr2R: 9,549,757– 9,550,931.

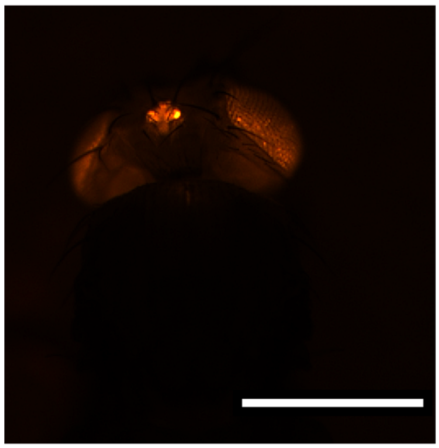

(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)
lncRNA:CR45312
CG17048
CG10814

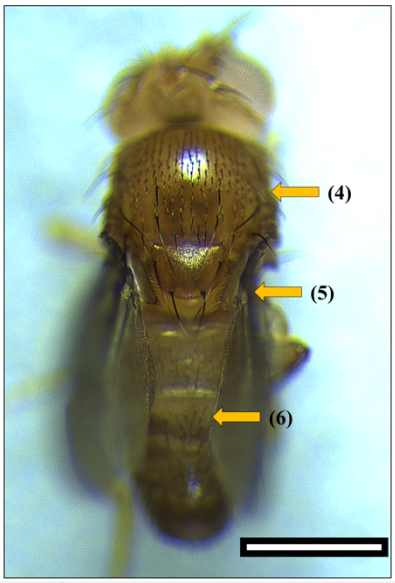
Supplementary Figure 14: lncRNA-15 (*RAMPAGEpeak_016800*) homozygous mutant. Homozygous null mutants for lncRNA-15 were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image inserts displayed eye phenotypes of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic loci. The strong expression of the RFP protein in the adult eyes can be detected easily (Scale bar, 0.5 mm). (d) Comparative profiling of lncRNA-15 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-15, the highest expression level of the lncRNA transcript in *Drosophila melanogaster* was detected at around 15 - 16 hr after egg laying. (e) Genomic loci of lncRNA-15 as visualised on the UCSC genome browser. Transcripts that have been annotated on Flybase were shown with dark blue bars. *CG17048* and *CG10814* are two protein-coding genes that are the closest to the lncRNA transcript. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-15 is expressed from the negative strand of the DNA and the transcript has been annotated on Flybase as *lncRNA:CR45312* with one isoform. A pair of gRNAs have been used to replace the promoter region of the lncRNA with a 3xP3-RFP cassette as shown (orange bar, blue highlight). The region excised is from chr2R: 13,362,387– 13,363,174.

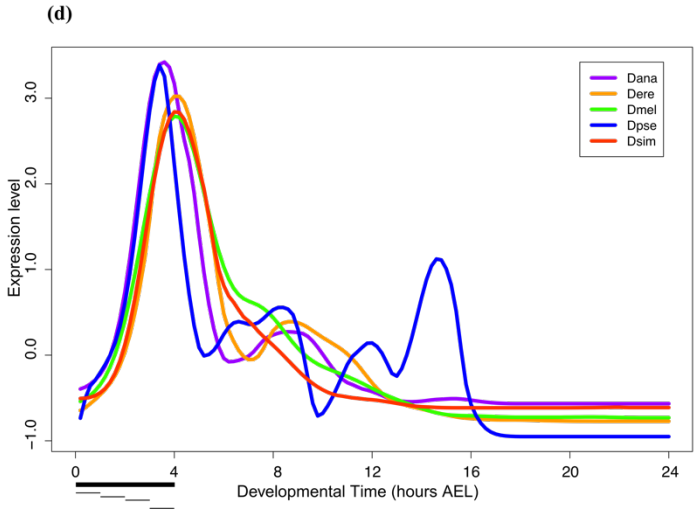

(a)
(1)
(2)
(3)
(b)
(4)
(5)
(6)
(c)
(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
00.
1.0
2.0
3.0
0
4
8
12
16
20
24
Developmental Time (hours AEL)
(e)
RAMPAGE (+)
RAMPAGE (-)
26,695,000
26,700,000
26,705,000
26,710,000
dm6
amon

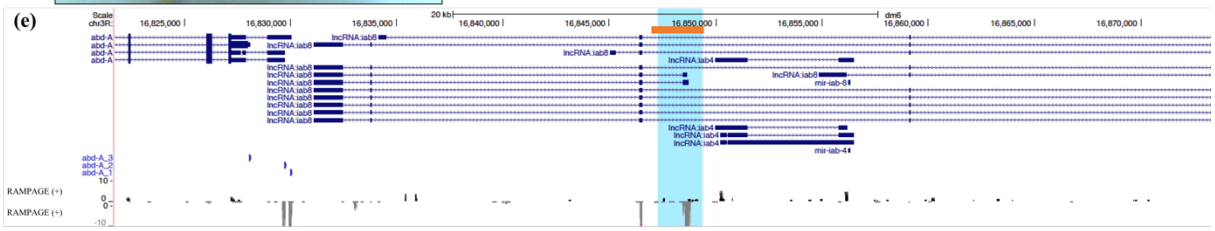
Supplementary Figure 15: lncRNA-16 (*RAMPAGEpeak_018802*) homozygous mutant.
Homozygous null mutants for lncRNA-16 were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic locus (Scale bar, 0.5 mm). (d) Comparative profiling of lncRNA-16 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-16, a clear and distinct peak was observed and the highest expression level of the lncRNA transcript in *Drosophila melanogaster* was detected at around 15 - 16 hr after egg laying. (e) Genomic locus of lncRNA-16 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars. *TweedleQ (TwdlQ)* is located immediately upstream of the lncRNA transcript whereas *amontillado (amon)* is downstream. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-16 is expressed from the plus strand of the DNA and the transcript has not been annotated on Flybase or the Ensembl genome assembly. A pair of gRNAs was used to replace the promoter region of the lncRNA with a 3xP3-RFP cassette as shown (orange bar, blue highlight). The region excised is from chr3R: 26,693,831– 26,694,688.


(a)
(1)
(2)
(3)


(c)


(b)
(4)
(5)
(6)


(d)
Dana
Dere
Dmel
Dpse
Dsim
Expression level
3.0
2.0
1.0
0.0
−1.0
0
4
8
12
16
20
24
Developmental Time (hours AEL)


(e)
Scale
chr3R:
16,825,000
16,830,000
16,835,000
20 kb
16,840,000
16,845,000
16,850,000
16,855,000
dm6
16,860,000
16,865,000
16,870,000
abd-A
lncRNA:iab8
lncRNA:iab4
mir-iab-8
mir-iab-4
abd-A_3
abd-A_2
abd-A_1
RAMPAGE (+)
RAMPAGE (+)

Supplementary Figure 16: lncRNA-19 (*FBgn0264857*) homozygous mutant.
LncRNA-19 homozygous mutants were viable (a,b) with normal anatomical features compared to wildtype (supplementary figure 4). A total of 10 homozygous males and 10 homozygous females were examined. (a) Lateral view of the adult male. (b) Dorsal view of the adult male. (c) Image shows the eye phenotype of flies with successful integration of 3xP3-RFP cassette at the lncRNA genomic loci. (Scale bar, 0.5 mm) (d) Comparative profiling of lncRNA-19 promoter activity during embryonic development. Embryos from five *Drosophila* species were collected and the libraries were generated using the RAMPAGE protocol. For lncRNA-19, the highest expression level of the lncRNA in *Drosophila melanogaster* detected at 5 - 6 hr after egg laying (AEL). (e) Genomic loci of lncRNA-19 as visualised on the UCSC genome browser. Transcripts annotated on Flybase are shown with dark blue bars with *lncRNA:iab4* and *abd-A* as the two closest genes to the lncRNA. From the RAMPAGE tracks, single nucleotide resolution of the TSS can be obtained (gray bar). The top RAMPAGE track shows promoter activity from the plus strand of the DNA, whereas the bottom track shows the opposite strand. LncRNA-19 is expressed from the negative strand of the DNA and is denoted in FlyBase as *lncRNA:iab8*. There are 13 annotated transcripts for this lncRNA. A pair of gRNAs was used to excise the promoter region of the lncRNA as shown (orange bar, blue highlight). The region excised is from chr3R: 16,847,383 – 16,849,389.

Chapter 4 |

A role for lncRNA-9 in the *Drosophila* nervous system

For the purpose of this thesis, I will outline my specific contributions to this project. To perform the RNA-fish experiments, the design and production of the probes for the RNA-fish were carried out by following companies with the lncRNA sequences provided by me. The Hybridization Chain Reaction (HCR) probes were ordered from Molecular Instrument, inc and the probes using the RNAscope® technology were made by Advanced Cell Diagnostics (ACD). The embedding, sectioning, staining and imaging of the embryos with the RNAscope® oligo probe set were performed by the staff from the CRUK CI Histopathology/ISH Core Facility (Bev Wilson and Dr. Julia Jones). Figure 26a was done by Dr Julia Jones. For the RNA-sequencing experiment, sequencing of the libraries was done by staff from the genomics core facility with the libraries prepared by me. The mapping of the modencode data and analyses of the RNA-seq data were performed by colleagues at our in-house bioinformatics core facility team (Dr. Ashley Sawle and Dr. Abigail Edwards). The rest of the figures presented here are the result of my own work.

4.1 Introduction

Cellular signalling and coordination from the embryo to a fully-grown adult are vital for the proper development of an organism. Different gene expression pathways have to be tightly regulated in time and space for the generation of different cell types and any mis-regulation would be deleterious for the system (Lee and Young, 2013; Peter and Davidson, 2011). During *Drosophila* embryonic development, the fertilised egg forms the future body axes and rapidly generates multiple cell lineages within 24 hr before the first instar larva emerges. This remarkable feat and speed thus require a well-coordinated system and logistics in place.

In recent years, ncRNAs were shown to have an important role in regulating the *Drosophila* genome and the majority of the studies looked at how miRNAs were regulating

gene expression during embryogenesis (Chawla and Sokol, 2011; Aboobaker *et al.,* 2005; Dai *et al.,* 2012). While the general importance of microRNAs in regulating various aspects of development has been well-established, relatively little was known about how lncRNAs fit in the picture and their functions remain to be addressed. In the previous chapter, I have made used of RAMPAGE data that was generated from all stages of *Drosophila* embryogenesis to identify potential lncRNA candidates. The molecular impact of knocking out the lncRNA loci was then studied. From the list of potential lncRNAs, 22 that had conserved expression across five *Drosophila* species were used for functional studies. Out of the 22 candidates, one was omitted as it was later annotated as a protein-coding gene and for 13 candidates lncRNA mutant flies were successfully generated used CRISPR/Cas9. A series of assays were carried out to assess the phenotypic consequences of candidate lncRNA loss, and I found two mutants, lncRNA-3 and lncRNA-9, with homozygous lethal phenotype (Chapter 3).

To better understand the role of *Drosophila* lncRNAs, I have chosen to focus on studying lncRNA-9 in greater depth in this chapter. To this date, thousands of lncRNAs have been identified in different organisms and many of these transcripts were found to be expressed in different tissues or were mis-regulated in many diseases (Xue *et al.,* 2015; Sanchez *et al.,* 2018; Huarte, 2015). Several bioinformatics analysis platforms were developed over the years to accommodate the increasing amount of lncRNA-related data and these publicly available datasets served as a good starting point for the characterisation of lncRNAs (Park *et al.,* 2014; Zhao *et al.,* 2016; The RNAcentral Consortium). However, these information does not provide any form of functional visualisation of the lncRNA within the cell nor does it tell us about the impact of loss of the transcript. Given that lncRNAs have been shown to be involved in many biological processes, it is imperative to invest efforts to decipher their roles and understand how these molecules contribute to development by further elucidating their mechanism of action.

DNA and RNA in situ hybridisation methods has been used extensively in molecular studies to analyse mRNA transcripts and some of them have the potential to be used as biomarkers in clinical settings (Schulz *et al.,* 2017; Lee *et al.,* 2014). Many lncRNA-related studies showed that these transcripts had tissue-specific or even cell-specific expression (Cabili *et al.,* 2015; Bindu *et al.,* 2018). RNA fluorescence in situ hybridization (RNA-FISH) has been the method of choice to study the localisation and expression of these often lowly abundant transcripts within the cell. From the information derived from the subcellular localisation of lncRNAs, we would have a better understanding of their expression dynamics

during embryonic development. The visualisation of their subcellular localisation could also provide insight into their physiological roles and allow researchers to carry out further investigations using different assays. Compared to mRNA for protein-coding genes, there remained a challenge to probe for lncRNA molecules using *in situ* hybridization (ISH) methods given their low abundance, poorly understood transcript structure and temporal expression within the cell. In order to solve this issue, several different ISH technologies have been developed throughout the years that greatly improved the visualisation of lowly abundant RNA transcripts.

In this study, RNA-FISH was used to understand the expression pattern and tissue-specificity of lncRNA-9 within the developing embryo and also probe for the compartmental location of the transcript within the cell. This was further paired with immunohistochemistry (IHC) to look for colocalization of the lncRNA transcript with proteins to determine in which cells the lncRNA was expressed within the tissue. Here, we identified that lncRNA-9 is expressed in the ventral nerve cord during *Drosophila* embryo development. This expression pattern coincides with a *Drosophila* neuron-specific marker *ELAV* (embryonic lethal abnormal vision) instead of *REPO* (reversed polarity), a marker that identifies glial cells.

The RNA-FISH experiment and the RAMPAGE data revealed a temporally dynamic and spatially restricted expression profile of lncRNA-9 implying a tightly controlled transcriptional network. To investigate the molecular phenotype induced by lncRNA-9 loss and the genes that were affected in the mutants, I generated RNA-seq libraries from wildtype and homozygous null mutant embryos at hourly intervals when lncRNA-9 was expressed to compare the transcriptomic profile of lncRNA mutants and wildtype embryos. From the differential gene expression analysis, clusters of genes were identified to be significantly downregulated in the mutant background. I have made use of pathway enrichment and GO (Gene ontology) analysis to look for downstream genes that were affected by the loss of lncRNA-9. A significant enrichment of GO terms related to nervous system development and function was identified. The results generated thus far have underlined a potential role for lncRNA-9 in influencing the development of the nervous system during embryogenesis and provide a good basis for other follow-up studies to further characterise its function.

4.2 Results

4.2.1 Supporting evidence for lncRNA-3 and lncRNA-9 transcripts

A number of high-quality and in-depth transcriptomic data pertaining to the development of *Drosophila* and *C. elegans* have been generated over the years by various groups within the modENCODE consortium. One of the goals of the modENCODE project was to characterise the functional elements in the *Drosophila* genome and since then, a wealth of data was gathered from thousands of experiments including RAMPAGE, ChIP-seq, and RNA-seq (Boley *et al.,* 2014; Brown and Celniker, 2015). In the previous chapter, two lncRNAs were found to be essential for *Drosophila* development but little was known about their function. Using available data from CAGE-seq (chapter 3), both transcripts were found to be expressed in the adult head. However, this information does not provide any other additional evidence about the localisation of the lncRNA and the tissue and cell types they were expressed in. Knowing the gene expression profile of the lncRNA transcript in the different tissues will greatly enhance our understanding of the possible role that is played by the lncRNA itself.

Using the GBrowse2 tool available on FLYBASE, I have mined for information pertaining to the expression of the two lncRNA transcripts with the help of the TSS information generated from RAMPAGE as a reference. To understand the characteristics of lncRNA-3 and lncRNA-9 genomic loci, I have mined existing data for chromatin signatures associated with the lncRNA transcripts. Active chromatin signatures such as H3K4me3, H3K36me3 and RNA Pol II occupancy have been previously used to identify transcribed regions (Shen *et al.,* 2012; Ardehali *et al.,* 2011; Huang *et al.,* 2018). In addition, other histone marks that define the chromatin landscape can also be found on the GBrowse2. Instead of chromatin marks for gene expression, the mono-methylated histone H3 lysine 4 (H3K4me1) mark was observed to occupy several regions throughout lncRNA-3 genomic loci. H3K4me1 is a chromatin signature found on inactive enhancers poised for activation during *Drosophila* development (Rada-Iglesias *et al.,* 2011; Bonn *et al.,* 2012; Koenecke et al., 2017). In addition, H3K27 methylation, a modification that is associated with gene repression was found to be deposited along the lncRNA transcript. The spread of both the enhancer-associated marker (H3K4me1) and a repression marker (H3K27me) from the gene body into the upstream promoter region of the lncRNA is suggestive that the expression of the lncRNA is likely to be tightly regulated. For lncRNA-9, only the recruitment of RNA Pol II just upstream from the TSS of the lncRNA was found. Neither H3K4me3 nor H3K36me3

chromatin marks were found to be associated with the genomic locus of lncRNA-9. Although an absence of chromatin signatures associated with active transcription (H3K4me3, H3K36me3 and Pol II) was observed for lncRNA-9, similar results were reported for *Drosophila* lncRNAs in the study by Chen *et al.,* 2016. The authors observed that a large percentage of *Drosophila* lncRNA genomic loci were not occupied by chromatin signatures. While the histone marks may provide useful information about the chromatin state, further studies will be needed to understand which features are better suited for understanding the existence of a lncRNA and its expression. Furthermore, as lncRNA is only expressed in a few cells, it will be hard to detect chromatin mark signals in ChIP experiments as they are from whole embryos and not sorted cells.

Besides the information gathered from ChIP-seq, the RNA sequencing data collected from various *Drosophila* tissues and developmental stages by the modENCODE consortium have provided an unparallel wealth of information for exploration of gene function. I have searched for the genomic loci of both lncRNA transcripts and found that both lncRNAs were expressed in the adult head, which is in line with the information that I have obtained from the CAGE-seq data. Furthermore, both lncRNA transcripts have been found to be expressed in the central nervous system with five different isoforms constructed from RNA-seq data for lncRNA-3 (Figure 21). Based on the PhyloCSF analysis, both lncRNAs are unlikely coding for functional peptides. Within the nervous system, lncRNA-9 had a modest expression in the neuroblast and neurons compared to lncRNA-3 (Figure 22). As lncRNA-9 was not annotated on Flybase, tissue specific expression data was only downloaded for lncRNA-3 (Table 17). The transcription factor (TF) HOT spots (high occupancy sites) track was generated to identify highly occupied target (HOT) regions using ChIP-seq data of 41 factors in embryos. Interestingly, the lncRNA-3 TSS and genomic loci was also a hotspot for transcription factors compared to lncRNA-9 and this may explain the rationale for its higher expression levels captured using RNA-seq and RAMPAGE. The expression of both lncRNA-3 and lncRNA-9 in the *Drosophila* central nervous system does not come as a surprise. Thousands of lncRNA transcripts have previously been reported to be pervasive in specific tissues with a large fraction expressed in the head and brain (Mercer *et al.,* 2008; Briggs *et al.,* 2015; Schor *et al.,* 2018).

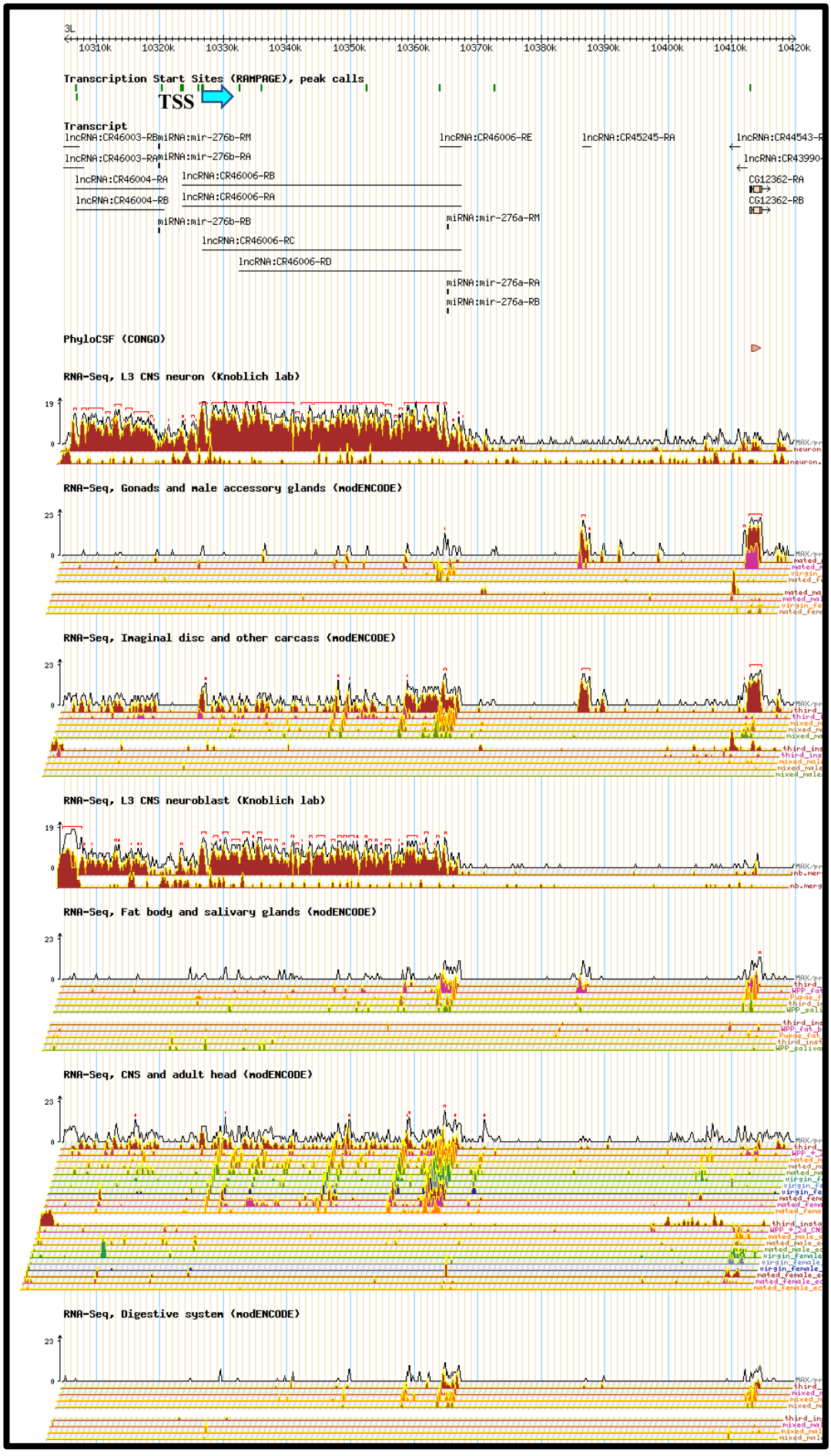

3L
10310k
10320k
10330k
10340k
10350k
10360k
10370k
10380k
10390k
10400k
10410k
10420k
Transcription Start Sites (RAMPAGE), peak calls
TSS
Transcript
lncRNA:CR46003-RB
lncRNA:CR46003-RA
lncRNA:CR46004-RA
lncRNA:CR46004-RB
miRNA:mir-276b-RM
miRNA:mir-276b-RA
lncRNA:CR46006-RB
lncRNA:CR46006-RA
miRNA:mir-276b-RB
lncRNA:CR46006-RC
lncRNA:CR46006-RD
lncRNA:CR46006-RE
lncRNA:CR45245-RA
lncRNA:CR44543-R
lncRNA:CR43990-
CG12362-RA
CG12362-RB
miRNA:mir-276a-RM
miRNA:mir-276a-RA
miRNA:mir-276a-RB
PhyloCSF (CONGO)
RNA-Seq, L3 CNS neuron (Knoblich lab)
RNA-Seq, Gonads and male accessory glands (modENCODE)
RNA-Seq, Imaginal disc and other carcass (modENCODE)
RNA-Seq, L3 CNS neuroblast (Knoblich lab)
RNA-Seq, Fat body and salivary glands (modENCODE)
RNA-Seq, CNS and adult head (modENCODE)
RNA-Seq, Digestive system (modENCODE)

Figure 21: A Gbrowse2 view of lncRNA-3.
Shown is a schematic screenshot of the *Drosophila* chromosome 3L (10,305,000 - 10,420,000) on the genome browser on FlyBase. Information on the directionality of expression for each of the transcript can be seen from the two sets of tracks provided. The top set of tracks refers to the positive strand whereas the bottom track denotes the minus strand. Using the data provided by the modENCODE Consortium, the full-length transcript for lncRNA-3 was identified with the help from the TSS information for this ncRNA (chr3L:10326628-10326781) previously downloaded from the RAMPAGE data. LncRNA-3 is annotated as *lncRNA:CR46006* on FlyBase with five isoforms (lncRNA:CR46006-RA/B/C/D/E) and it is expressed from the positive DNA strand. All of the isoforms contain two exons that are separated by an intronic region except for lncRNA:CR46006-RE (not shown). In the browser, each TSS captured by RAMPAGE is indicated by a green bar/arrow under "Transcriptional start site (RAMPAGE), peak calls". For the convenience of identifying the TSS, a bigger cyan arrow has been drawn to indicate the TSS of the lncRNA. *MiR-276b* and *lncRNA:CR46004* are located upstream of lncRNA whereas *miR-276a* is localised downstream of the lncRNA-3 transcript. A multi-species nucleotide sequence alignment tool used to denote protein-coding regions is represented by the PhyloCSF track. Protein-coding regions are represented with a red bar/arrow as seen for *CG12362*. The Phylocsf track for lncRNA-3 indicates that no open reading frame has been detected, further demonstrating that this transcript is a ncRNA. The modENCODE tissue-specific stranded RNA-seq data contains a dynamic genome-wide expression profiling of *Drosophila* (Brown *et al.,* 2014). RNA-seq data have been plotted from different tissue types including digestive system, fat body and salivary glands, imaginal disc and other carcass, CNS and adult head, gonads and male accessory glands, L3 CNS neuron and L3 CNS neuroblast (log2 scale). From the coverage profiles, lncRNA-3 was found to be highly expressed in the central nervous system tissues (3rd instar larvae CNS neuron, 3rd instar larvae CNS neuroblast and head of the adult). Transcription factor (TF) HOT spot analysis from whole embryo ChIP is not shown. For each of the tissue-specific RNA-seq profiles, the default log2 signal scaling method was applied.

.

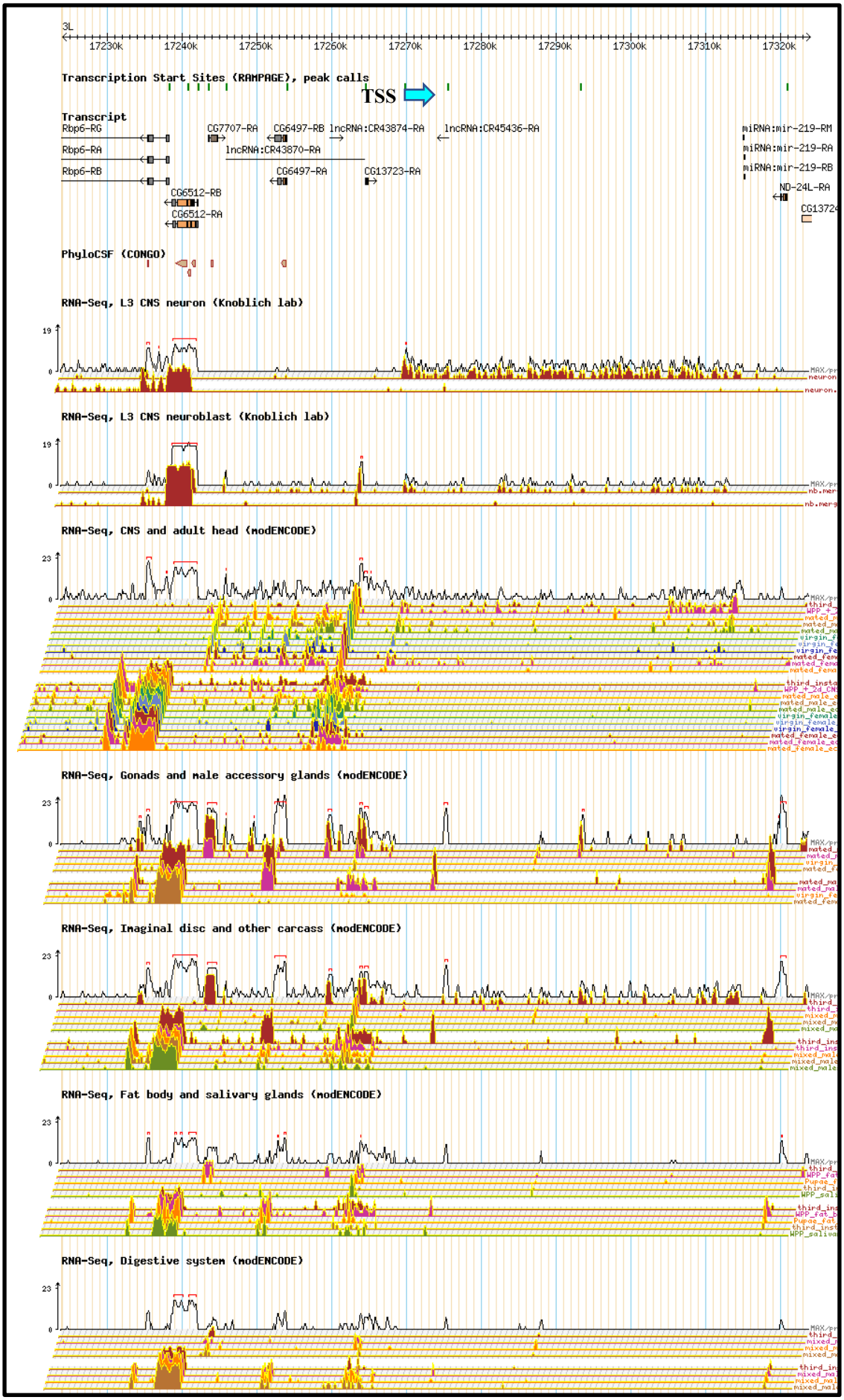

3L
17230k
17240k
17250k
17260k
17270k
17280k
17290k
17300k
17310k
17320k
Transcription Start Sites (RAMPAGE), peak calls
TSS
Transcript
Rbp6-RG
Rbp6-RA
Rbp6-RB
CG7707-RA
CG6497-RB
lncRNA:CR43874-RA
lncRNA:CR45436-RA
lncRNA:CR43870-RA
CG6497-RA
CG13723-RA
CG6512-RB
CG6512-RA
miRNA:mir-219-RM
miRNA:mir-219-RA
miRNA:mir-219-RB
ND-24L-RA
PhyloCSF (CONGO)
RNA-Seq, L3 CNS neuron (Knoblich lab)
RNA-Seq, L3 CNS neuroblast (Knoblich lab)
RNA-Seq, CNS and adult head (modENCODE)
RNA-Seq, Gonads and male accessory glands (modENCODE)
RNA-Seq, Imaginal disc and other carcass (modENCODE)
RNA-Seq, Fat body and salivary glands (modENCODE)
RNA-Seq, Digestive system (modENCODE)

Figure 22: A Gbrowse2 view of lncRNA-9
This diagram is a schematic screenshot of the genomic locus of lncRNA-9 on the *Drosophila* chromosome 3L (17,223,946 - 17,323,946). Information on the directionality of expression for each of the transcript can be seen from the two sets of tracks provided. The top set of tracks refers to the positive strand whereas the bottom track denotes the minus strand. Notice that the full-length transcript for lncRNA-9 is not annotated on FlyBase. The TSS for lncRNA (chr3L:17269741-17269766) is captured by RAMPAGE and it is indicated with a green arrow/bar under "Transcriptional start site (RAMPAGE), peak calls". A bigger arrow in cyan has been drawn to indicate the TSS of the lncRNA. LncRNA-9 is expressed from the positive strand of the DNA with *CG13723* located upstream of the transcript. LncRNA:CR45436 is expressed from the opposite direction and it is annotated downstream of the lncRNA. A multi-species nucleotide sequence alignment tool used to denote protein-coding regions is represented by the PhyloCSF track. Protein-coding regions are represented with a red bar/arrow as seen for *CG6497*. NONCODE 2016 was used to search for the lncRNA transcript and using Phylocsf analysis, the ncRNA did not appear to have any short open reading frames within the transcript that was constructed on NONCODE (data not shown). The Phylocsf track displayed here indicates that no open reading frame has been detected, further suggesting that this transcript is a ncRNA. Transcription factor (TF) HOT spot analysis from whole embryo ChIP is not shown. For each of the tissue-specific RNA-seq profiles, the default log2 signal scaling method was applied.

We next asked whether lncRNA-3 and lncRNA-9 carry out their function through a cis- or trans-acting mechanism. In order to answer this question, rescue transgenic models were created using clones from Drosophila melanogaster BAC libraries (Venken *et al.,* 2009). For the rescue experiments, the transgenic flies generated for lncRNA-3 (CH322-18B08 and CH322-151D21) and lncRNA-9 (CH322-170N22 and CH322-122O7) are listed in Table 2.2. Each of these clones were between ~21 kb and 40 kb in size and overlapped with the deleted region of the corresponding lncRNA mutant (Figure 23). The transgenic rescue flies were established on the second chromosome and subsequently crossed to a double balancer (w[1118]; If/Cyo ; MKRS/TM6B, Tb[1]). The lncRNA mutants were crossed to the same double balancer fly and both stocks were subsequently mated to generate flies that only expressed the lncRNA transcript from the transgene. The number of adults of was scored to assess rescue of the mutant by the transgene. However, using these lines I was unable to revert the phenotype caused by both lncRNA-3 and lncRNA-9. These results suggest that both lncRNAs were unlikely to be regulating genes in a trans-acting manner. Following these analyses, while both lncRNA candidates have presented interesting phenotypes with CRISPR/Cas9 mutagenesis, I will be focusing on characterising lncRNA-9 in the next parts of this chapter as the phenotype of the lncRNA has been fully validated to be dependent on the lncRNA using alternative mutants. Additional supporting evidence will be needed to confirm the phenotype for lncRNA-3 is indeed depended on the lncRNA absence rather than confounding factors such as off-target effects or deletion of DNA sequences.

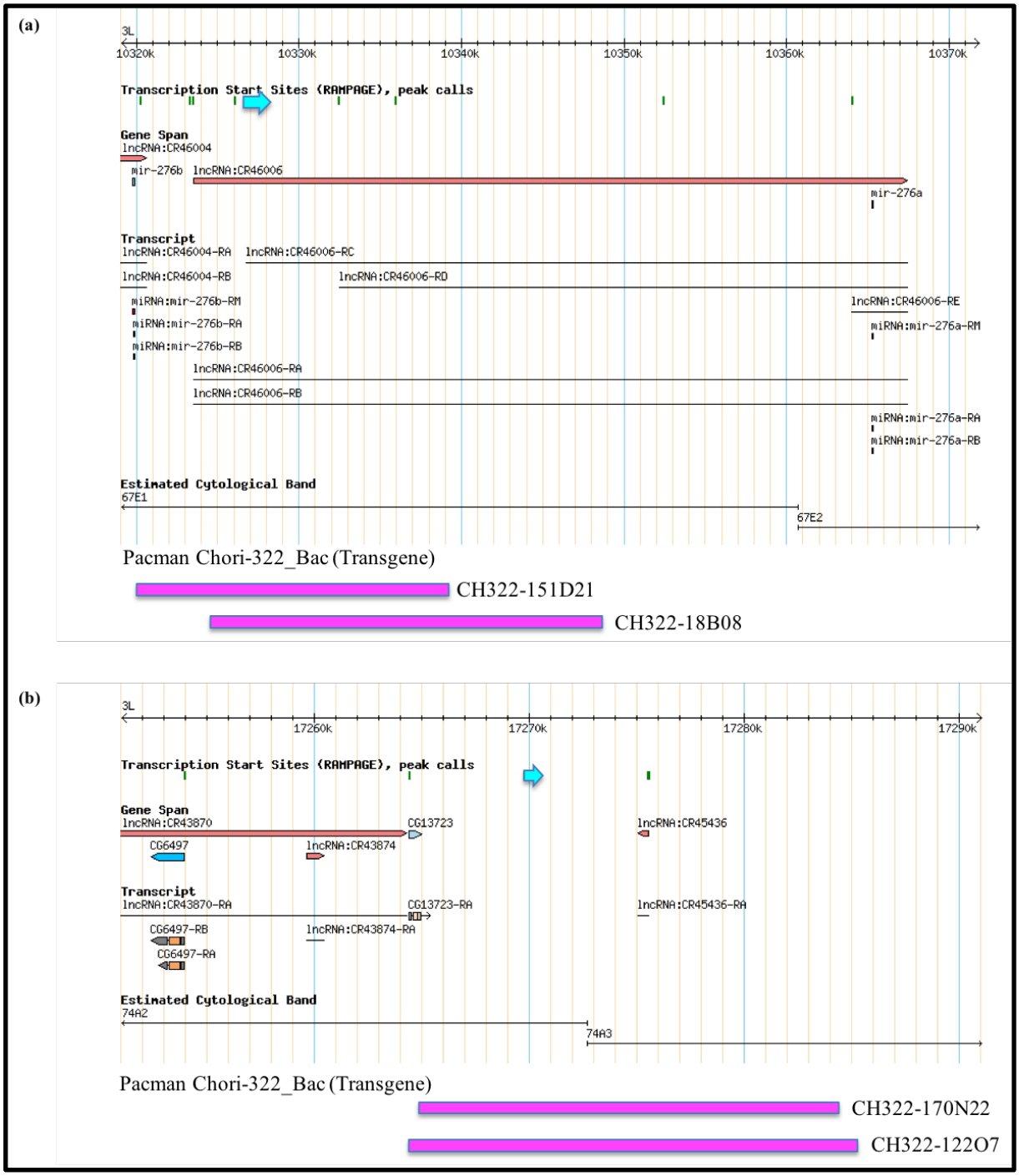

(a)
3L
10320k
10330k
10340k
10350k
10360k
10370k
Transcription Start Sites (RAMPAGE), peak calls
Gene Span
lncRNA:CR46004
mir-276b
lncRNA:CR46006
mir-276a
Transcript
lncRNA:CR46004-RA
lncRNA:CR46006-RC
lncRNA:CR46004-RB
lncRNA:CR46006-RD
miRNA:mir-276b-RM
lncRNA:CR46006-RE
miRNA:mir-276b-RA
miRNA:mir-276a-RM
miRNA:mir-276b-RB
lncRNA:CR46006-RA
lncRNA:CR46006-RB
miRNA:mir-276a-RA
miRNA:mir-276a-RB
Estimated Cytological Band
67E1
67E2
Pacman Chori-322_Bac (Transgene)
CH322-151D21
CH322-18B08
(b)
3L
17260k
17270k
17280k
17290k
Transcription Start Sites (RAMPAGE), peak calls
Gene Span
lncRNA:CR43870
CG13723
lncRNA:CR45436
CG6497
lncRNA:CR43874
Transcript
lncRNA:CR43870-RA
CG13723-RA
lncRNA:CR45436-RA
CG6497-RB
lncRNA:CR43874-RA
CG6497-RA
Estimated Cytological Band
74A2
74A3
Pacman Chori-322_Bac (Transgene)
CH322-170N22
CH322-122O7

Figure 23: BAC rescue for lncRNA-3 and lncRNA-9.
Schematic diagram of the genomic locus and the design of rescue lines for lncRNA-3 and lncRNA-9. BAC constructs containing the wild-type version of the deleted sequences (magenta bars) were used to reintroduce them. PACMAN Genome Browser and FlyBase gbrowse were used to search for the BAC Clones containing the lncRNA of interest. For each lncRNA, two clones were selected from the "CHORI-322" BAC library (Venken *et al.,* 2009). (a) Genomic loci of lncRNA-3 with the TSS represented by an arrow in cyan (chr3L:10326628-10326781). For lncRNA-3, the deleted region spans from chr3L: 10,325,557 – 10,327,321. The BAC clones CH322-18B08 (3L: 10,324,751-10,348,381) and CH322-151D21 (3L: 10,320,125-10,339,114) were selected and separately introduced to rescue the phenotype. (b) For lncRNA-9, the region excised is from chr3L: 17,269,407 – 17,270,407. The BAC clones CH322-170N22 (3L: 17,264,834-17,284,252) and CH322-122O7 (3L: 17,264,275-17,285,125) were found to be overlapping with the deleted region and used for the rescue experiment.

4.2.2 LncRNA-9, a nuclear RNA transcript expressed during *Drosophila* embryogenesis

Many studies have highlighted that lncRNAs present cell- and tissue-specific expression that would be the basis for their physiological functions (Wilk *et al.,* 2016; Cabili *et al.,* 2015). RNA-FISH was used to probe the general expression pattern and subcellular localisation of lncRNA-9. In this study, the RNAscope® oligo probe set was designed and obtained from Advanced Cell Diagnostics (ACD). On the other hand, in situ hybridization chain reaction (HCR) probes against lncRNA-9 were generated by Molecular Technologies. Oligo probe sets from RNAscope® were first obtained to stain *Drosophila* embryo sections due to the lack of protocols for working with whole-mount *Drosophila* embryos. Probes from Molecular Technologies were used much later for the staining of whole-mount embryos.

A chromogenic in situ hybridization assay was carried out using the RNAscope® oligo probe set on sectioned late stage *Drosophila* embryos (stage 16-17). The embedding, sectioning, staining and imaging of the embryos with the RNAscope® oligo probe set were performed by Bev Wilson and Dr. Julia Jones from the Histopathology/ISH Core Facility with the embryos prepared by me. The double Z probes were designed to bind to the lncRNA transcript in tandem with high specificity. Each Z probe consists of 2 components. The base region of the Z probe was made to be complementary to target sequence whereas the top end would bind to the pre-amplifier for subsequent signal amplification. Upon hybridisation with the target lncRNA sequence, the pre-amplifiers would interact with the top end of the double Z followed by the binding of amplifiers to the binding sites found on each pre-amplifier. Productive amplification only occurs when both Z probes are binding the target in tandem thereby reducing the background resulting from unspecific binding of each individual Z probe. Label probes containing chromogenic enzymes such as horseradish peroxidase were later added to allow conjugation with the 20 binding sites on each amplifier. The target lncRNA was visualised using a standard bright-field microscope after the addition of substrate (https://acdbio.com). To visualise the localisation of the lncRNA using DNA-HCR probes, fixed whole-mount embryos were prepared according to the protocol provided by the company (Choi *et al.,* 2016). Briefly, the technology consisted of three components: initiator probe, amplifier probe H1 and amplifier probe H2. The initiator probes consisted of two parts. One part of the initiator contains specific sequences for hybridising to regions of the lncRNA transcript. The other part of the initiator was made for initiating polymerisation of the two amplifier probes. Upon exposure to the unhybridized end of the initiator probe, a self-

assembling cascade hybridization reaction with the two fluorescently labelled amplifiers would eventually lead to a linear amplification of the signal (Choi *et al.,* 2016; Choi *et al.,* 2018). Using embryos at stage 16-17 of development, both methods revealed strikingly similar spatio-expression of the lncRNA. LncRNA-9 transcripts were localised to the the ventral nerve cord of the developing embryo (Figure 24, 25, 26). This localisation of lncRNA-9 to the central nervous system consistently with the earlier results from the modENCODE data. In addition, co-localisation of the transcript with DAPI staining indicated that the lncRNA is localised in the nucleus (Figure 24, 25, 26d), in agreement with a potential function in cis as suggested by the BAC rescue experiments.

4.2.3 LncRNA-9 is a neuron-specific lncRNA

The ventral nerve cord consists of a wide variety of cell types including neuroblasts, neurons and glial cells. To further characterise the lncRNA-9 transcript, we asked if it was localised in neuroblast/neuronal cells, glial cells or both. To visualise lncRNA expression in these cell types, I performed multiplexed RNA-FISH against this lncRNA and neurogenic gene markers expressed in stage 16-17 embryos. In this experiment, we have made use of the HCR protocol with antibodies as it would allow us to visualise co-localisation in a whole mount embryo. LncRNA-9 was found to be co-localised with the *ELAV* protein in the neuroblast/neurons and this result agreed with the modENCODE data in section 4.2.1 (Figure 24, 25 and Figure 26). However, we observed that the co-localization of lncRNA-9 and *ELAV* was restricted to only a subset of the neuronal cells. This finding would explain the low expression levels detected for this transcript in a bulk RNA-seq experiment and also reflects the point that lncRNAs have cell-specific expression within the pool of neuronal cell types generated.

Using this information, I hypothesized that in the homozygous null mutant, the neurons that expressed lncRNA-9 would not be created and we would see gaps or a much shorter form of the ventral nerve cord (VNC) in the mutant embryo. Interestingly, although the lncRNA was found at spots throughout the VNC, the phenotype of the mutant embryo did not match our expectation. Instead, the VNC was found to have retracted in the majority of the wildtype embryos at this stage whereas in the homozygous null mutants, the ventral nerve cord remained extended. This finding suggested that lncRNA-9 mutants could have been arrested

at stage 16 or even earlier. Alternatively, the loss of the lncRNA transcript could have affected the VNC condensation process, resulting in an embryo with an elongated ventral nerve cord. For the staining with the glial cell marker *repo,* we found that the lncRNA did not co-express with cells of the glial lineage (Figure 26c), indicating that it is restricted to the neuronal lineage.

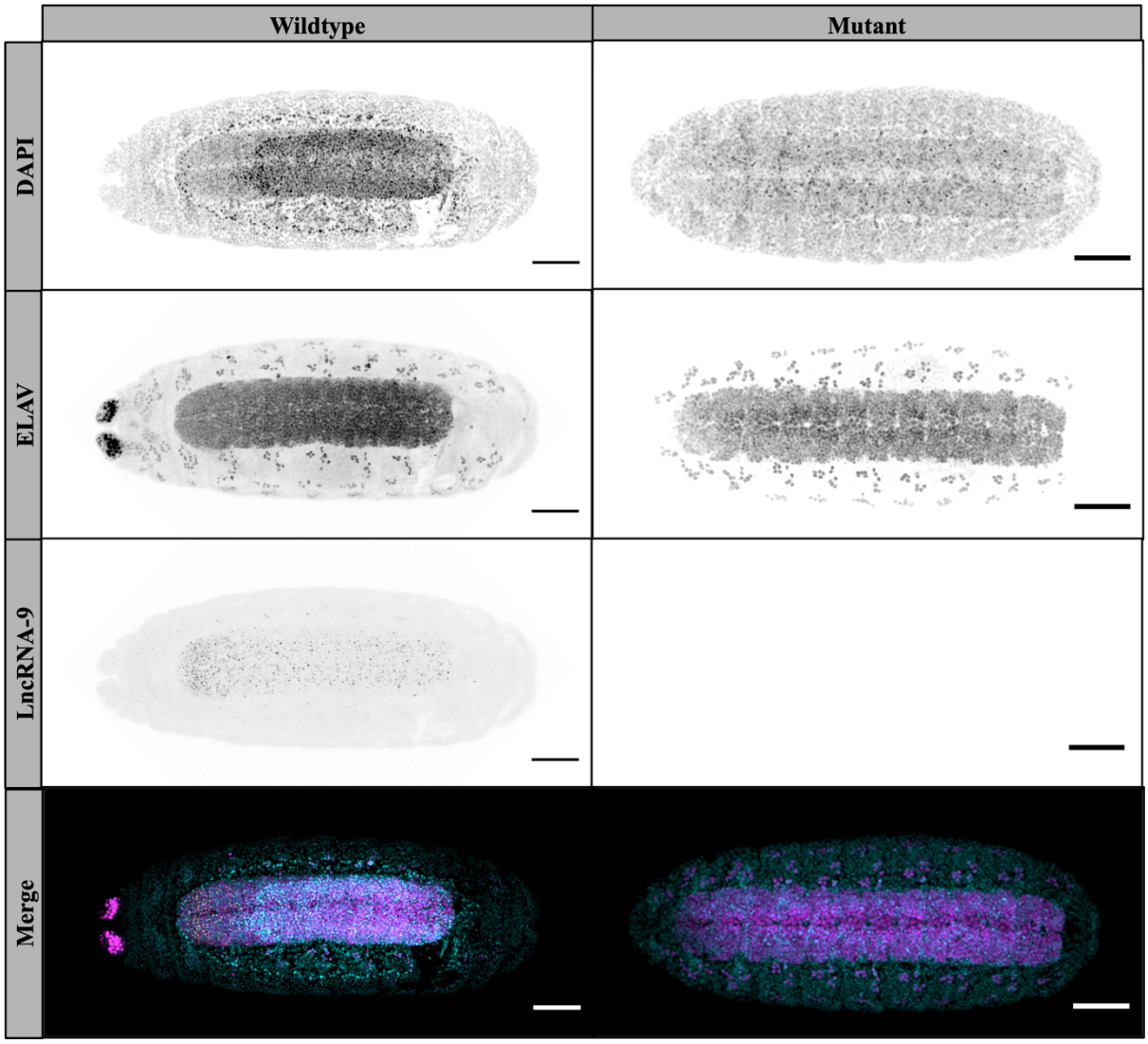


Figure 24: Spatiotemporal expression of lncRNA-9 transcript with *ELAV* (Dorsal view)
In situ hybridization of stage 16 embryos using a multiplexed RNA-FISH protocol. The lncRNA-9 transcript was found to be expressed in the VNC of the developing nervous system (Stage 16-17 embryos). Expression and localisation of lncRNA-9 and *ELAV* in the embryo are shown by immunofluorescence (DNA, cyan; *ELAV*, magenta; lncRNA-9, yellow). Both wildtype (left column) and lncRNA homozygous null mutant (right column) embryos were staged and collected for this experiment. In the homozygous null mutant generated with Cas9-HDR, I did not detect any lncRNA expression using the same settings on the microscope that was used for wildtype embryos. The VNC of the developing mutant embryo remain elongated compared to wildtype embryos of the similar stage. Images were taken with the anterior end on the left and the posterior side on the right. During the imaging of the embryos, 80-100 slices were taken for each embryo and a total of 20 embryos were analysed for each genotype. All scale bars represent 50 µm.

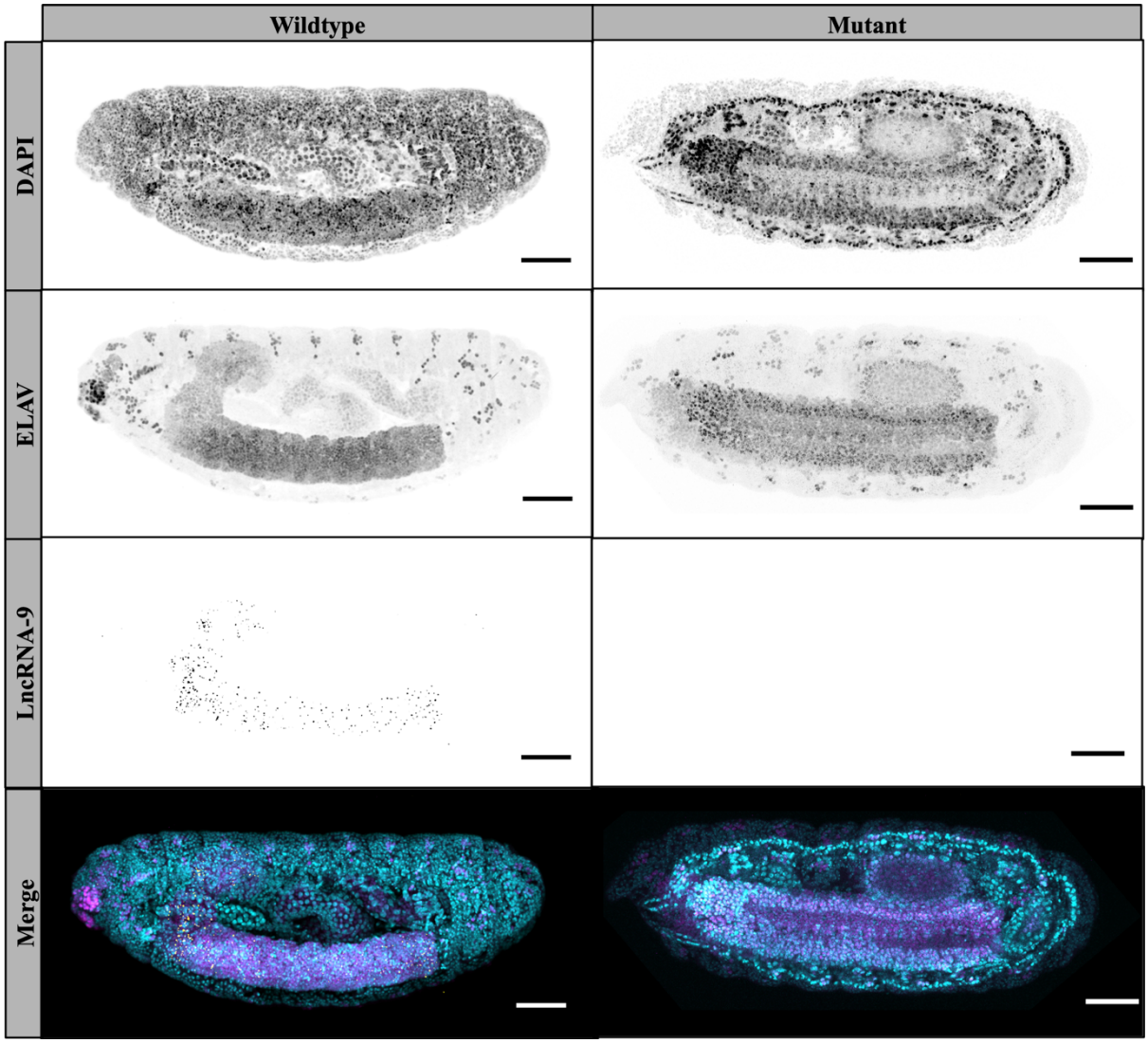


Figure 25: LncRNA-9 is expressed in the embryonic ventral nerve code (Lateral view)
Using a multiplexed RNA-FISH protocol, the lncRNA-9 transcript was found to be expressed in the VNC of the nervous system (Stage 16-17 embryos). Expression and localization of lncRNA-9 and *ELAV* in the embryo are shown by immunofluorescence (DNA, cyan; *ELAV*, magenta; lncRNA-9, yellow). Both wildtype (left column) and lncRNA homozygous null mutant (right column) embryos were staged and collected for this experiment. In the homozygous null mutant that have been generated with Cas9-HDR, I did not detect any lncRNA expression using the same microscope settings as for the wildtype embryos. Images were taken with the anterior end on the left and the posterior side on the right. During the imaging of the embryos, 80-100 slices were taken for each embryo and a total of 20 embryos were analysed for each genotype. All scale bars represent 50 µm.

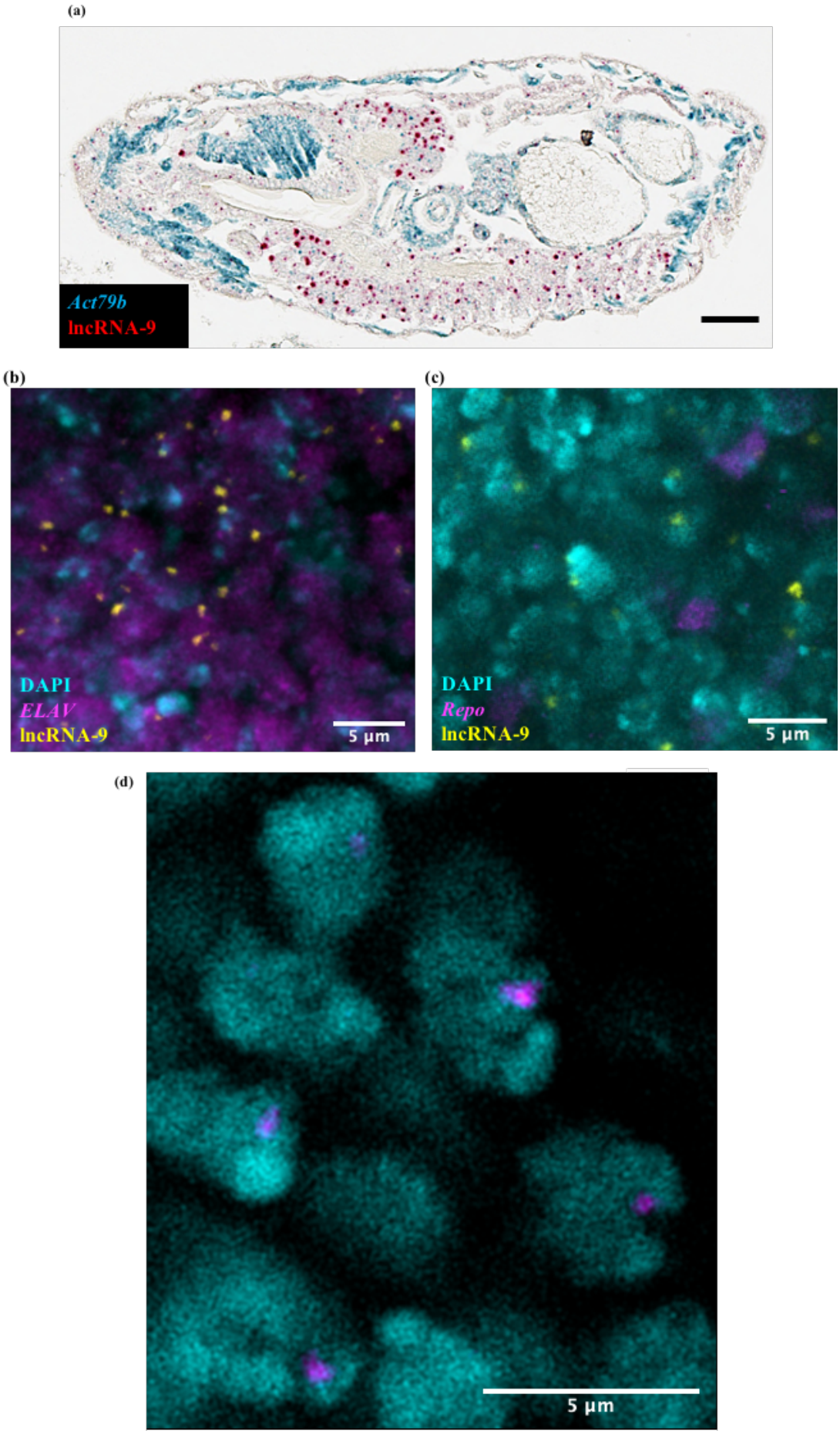

(a)
Act79b
lncRNA-9
(b)
DAPI
ELAV
lncRNA-9
5 µm
(c)
DAPI
Repo
lncRNA-9
5 µm
(d)
5 µm

Figure 26: Localisation of lncRNA-9 transcript in the developing embryo.
(a) Stage 16-17 wildtype embryos were stained with oligo probe sets for lncRNA-9 (brown) and act79b (blue). The signal from the RNA was visualised using horseradish peroxidase and depicted in brown. Similar to earlier results, lncRNA-9 transcripts were localised to the VNC of the developing embryo. Scale bar represent 30 µm. (b) The embryos were co-stained with the lncRNA smFISH probes (yellow) and *ELAV* (magenta) neuronal cell marker. A colocalization of the signals can be observed indicating that lncRNA-9 transcripts were expressed in neurons. (c) Co-staining of embryos with lncRNA smFISH probes (yellow) and glial cell marker, *Repo* (magenta). The lncRNA-9 transcript was not found to be co-expressed with *Repo*. Zoomed in images taken with a 63x lens (d) The embryos were co-stained with the lncRNA smFISH probes (yellow) and DAPI. Zoomed in images taken with a 100x magnification lens. LncRNA-9 was found to co-localise with the DAPI staining in the nucleus. In all 3 images (b,c,d), DAPI was used to stain for the nucleus (cyan). Unless otherwise indicated, the scale of the scale bar is as shown.

4.2.4 Differential expression analysis of lncRNA-9 mutant

To investigate the molecular consequences of lncRNA-9 knockout on downstream targets, I carried out RNA-Seq. For the transcriptome analysis, RNA-seq libraries were generated from wildtype and homozygous lncRNA-9 mutant embryos that were collected at hourly timepoints from 10-24 hr (14 timepoints). The term used to describe each timepoint collection is as follow: T10 (10 hr – 11 hr AEL); T11 (11 hr – 12 hr AEL); T12 (12 hr – 13 hr AEL); T13 (13 hr – 14 hr AEL); T14 (14 hr – 15 hr AEL); T15 (15 hr – 16 hr AEL); T16 (16 hr – 17 hr AEL); T17 (17 hr – 18 hr AEL); T18 (18 hr – 19 hr AEL); T19 (19 hr – 20 hr AEL); T20 (20 hr – 21 hr AEL); T21 (21 hr – 22 hr AEL); T22 (22 hr – 23 hr AEL) and T23 (23 hr – 24 hr AEL). For each mutant library, the embryos were manually collected by hand using the fluorescence balancer to distinguish them from their heterozygous siblings. The wildtype RNA-seq libraries were generated from control flies collected from a different cage. I collected 100 embryos for each library and a total of 56 libraries were generated (2 replicates per timepoint, 28 libraries per genotype). The sequenced reads were mapped to the *Drosophila* dm6 reference genome and alignment was carried out using HISAT2. For the purpose of differential gene expression analysis, all samples fulfilled the requirement of at least 30 million reads per library. Coverage plots for each sample were visualised in the UCSC genome browser (Figure 27). The TSS of the transcript was located using information from the RAMPAGE data (Figure 27a). In agreement with the phenotype validation performed by PCR of the genomic locus and the RNA-FISH results, the RNA-sequencing data from the mutant embryos showed almost a complete lack of coverage for the genomic region of lncRNA-9 (Figure 27c) compared to wildtype embryos (Figure 27b).

In this study, we looked for genes in the RNA-seq libraries that were differentially expressed when compared to wildtype embryos. We reasoned that because of the broad developmental defects observed, the main pathways affected would be differentially regulated across several stages. The first approach was to focus on the top 200 genes with the most significant changes based on their adjusted *p*-value (padj) (Supplementary Table 1.1-1.14). For each time point, I have subjected the gene cohorts to PANTHER Overrepresentation Test (Released 20200407) via The Gene Ontology (GO) Resource (The Gene Ontology Consortium, 2019). The annotation data sets used include PANTHER GO-Slim Biological Process, PANTHER GO-Slim Molecular Function and ANTHER GO-Slim Cellular Component (PANTHER version 15.0 Released 2020-02-14). During the GO enrichment analysis, only results with a false discovery rate [FDR] of less than 0.05 were used. The results from the GO analysis has been arranged in a “Hierarchical” fashion with the most

specific subclass indented directly above its parent term. GO terms that are over-represented or enriched will be selected for further studies.

With reference to the RAMPAGE data for lncRNA-9 (Chapter 3, supplementary figure 8), there are a couple of key timepoints during the expression of the lncRNA. The lncRNA is expressed prior to T10 (corresponding to 10 hr – 11 hr AEL) and increases in expression level thereafter. This is followed by a peak in expression level at around T16 (corresponding to 16 hr – 17 hr AEL). However, the samples were collected from T10 onward as the fluorescence signal is only detected in the embryos after this time due to the promoter of the balancer. From the results obtained (Table 18), GO term enrichment showed significant enrichment in neuronal processes (synaptic signalling, chemical synaptic transmission, synaptic transmission, glutamatergic) at 2 timepoints, T10 and T17. At T10, genes involved in ammonium transport, three of them (*Syt12, Syt1, Amt*) are found in our list (22.51 enrichment). Genes involved in synaptic transmission include *Syt12, Syt1, mGluR, nAChRalpha7, nAChRalpha1, Grik, SecCl and GluRIIE*. While the expression of lncRNA-9 starts to increase after T12, the TSS can be detected as early as T10 and peaks at T16. It is not surprising that biological processes associated with the nervous system were enriched at T17 as this is the timepoint right after the highest expression of the lncRNA transcript itself. Besides an enrichment for terms related to the nervous system, biological processes pertaining to the DNA repair pathway were detected (double-strand break repair, metabolic process, cellular response to DNA damage stimulus).

In order to examine the significantly up- and down-regulated genes, the gene lists were further sorted into two categories based on their log fold-change value (down- or upregulated) and then ranked based on their adjusted *p*-value (padj). The lists of genes for each timepoint used for the GO enrichment analysis can be found in Supplementary Table 2.1-2.14 (downregulated genes) and Supplementary Table 3.1-3.14 (upregulated genes). As the expression of lncRNA-9 increases from timepoints T13 (corresponding to 13 hr – 14 hr AEL) and peaks at T16 (corresponding to 16 hr – 17 hr AEL), I first examined the genes that are differentially expressed at T16 as well as the two timepoints later (T17 and T18). From the analysis, GO term enrichment of downregulated genes revealed significant enrichment in biological processes related to neuronal function such as chemical synaptic transmission activity (synaptic signalling) and modulation of trans-synaptic signalling in all three timepoints (Table 19.1- 19.2). The molecular functions of these genes were associated with roles in transmitter-gated ion channel activity (neurotransmitter receptor activity, extracellular

ligand-gated ion channel activity), anion transmembrane transporter activity, inorganic cation transmembrane transporter activity and active transmembrane transporter activity (Figure 28). Further analysis of the cellular components revealed categories associated with neuronal function (neuron projection, synapse, postsynaptic membrane). Genes involved in synaptic transmission include *Syt12, mAChR-B, DAT, nAChRalpha7, nAChRalpha2, Grik, and SecCl.*

Besides examining the immediate effects after lncRNA-9 expression reaches its peak, I also analysed the genes that are differentially expressed at earlier time points (T13 -T15). Just prior to the peak expression of lncRNA-9, the genes that are downregulated are enriched for small molecule catabolic processes (carboxylic acid catabolic process, branched chain metabolic process). However, the results observed for both molecular functions and cellular components domains for most of the timepoints were not statistically significant. While the metabolism of amino acids is essential during *Drosophila* embryogenesis and distinct roles are played by different amino acids in different stages during embryogenesis, little is known regarding the role of branched chain amino acid catabolism and their contribution to nervous system development (An *et al.,* 2014; Esslinger *et al.,* 2013). On the other hand, loss of lncRNA-9 appeared to have led to the activation and upregulation of genes associated with many terms related to cell cycle related activities such as DNA metabolism and DNA repair mechanisms (Table 20.1-20.3).

Characteristics of differentially expressed genes using GO pathway analysis

While the pathways enriched among the upregulated genes warrant further investigation, I have chosen to focus on the genes that were downregulated as the outcomes of the analysis were in line with results shown in the RNA-FISH data and further authenticate a role of lncRNA-9 in neural function. The analysis of Reactome (version 65 Released 2019-12-22) pathways enrichment was evaluated using the dataset available by the GO consortium. Using genes from T16-T18, pathways enriched for components in these lists include classes such as ‘neuronal system’ and ‘transmission across chemical synapses’. Only results with an FDR of less than 0.05 is used. The genes identified under the ‘transmission across chemical synapses’ class include *prt, nSyb, Nmdar1, Ggamma30A, CG44098, Ir76b, SecCl and Hsc70-2*. Factors that have been identified in the class of ‘neuronal system’ contain all the genes mentioned earlier and *CG18095*.

LncRNAs are known to regulate mRNAs in cis and it is believed that their functions are related to their neighbouring protein-coding genes. Since I was unable to rescue the phenotype of lncRNA-9 using a BAC rescue experiment, this result suggested a possible cis-acting role. Here, I have examined the expression of neighbouring mRNAs in the 200 kbp of sequence representing the region closest to the identified lncRNAs (Figure 29). From the RNA-sequencing analysis, 26 genes are found to be within 200 kbp from the TSS of the lncRNA. Target of Poxn (*Tap*) and Secretory Chloride Channel (*SecCl, CG7589*) are two of the genes that are located downstream of lncRNA-9 and their expression was found to be downregulated in the lncRNA mutant (padj<0.05, $\log_2$FC>1).

SecCl is a cys-loop glycine-gated chloride ion channel that is 200 kbp from the TSS of lncRNA-9 and has also been identified from the GO pathway analysis previously. The only known function of this protein is that it is a mediator of fluid secretion and the molecular phenotype associated with *SecCl* has been annotated on FlyBase as partially lethal and increased mortality during development. It is involved in the biological processes negative regulation of neuronal action potential; chemical synaptic transmission and regulation of circadian rhythm (Frenkel *et al.,* 2017). Proteins of the cys-loop glycine-gated chloride ion channel have previously been reported to be expressed in the nervous system and in muscle cells (Sine and Engel, 2006). They act as post-synaptic communicators and interact with neurotransmitters released from the pre-synapse (Sine and Engel, 2006). As little is known about the interactors with *SecCl*, I searched FlyBase for other glycine receptors and found *CG7446* (*Grd*), *CG17336* (*Lcch3*), *CG12344* and *CG10537* (*Rdl*). From the RNA-seq data, a significant downregulation was observed for *CG10537* (*Rdl*) and *CG12344* immediately after the peak of the lncRNA whereas both *Lcch3* and *Grd* are downregulated from T14-T18 in the lncRNA-9 mutant (padj<0.05, $\log_2$FC>1). On the other hand, Tap is a basic helix-loop-helix (bHLH) neural transcription factor that is the homolog for Neurogenin/NeuroD. It is located 92 kbp downstream from the TSS of lncRNA-9 and while there is little experimental evidence available for Tap, the molecular phenotype for flies lacking Tap is closely similar that observed in lncRNA-9 mutants. Even though there are a few escapers to early larval stages, the majority of the Tap mutants were reported as embryonic lethal (Yuan *et al.,* 2016).

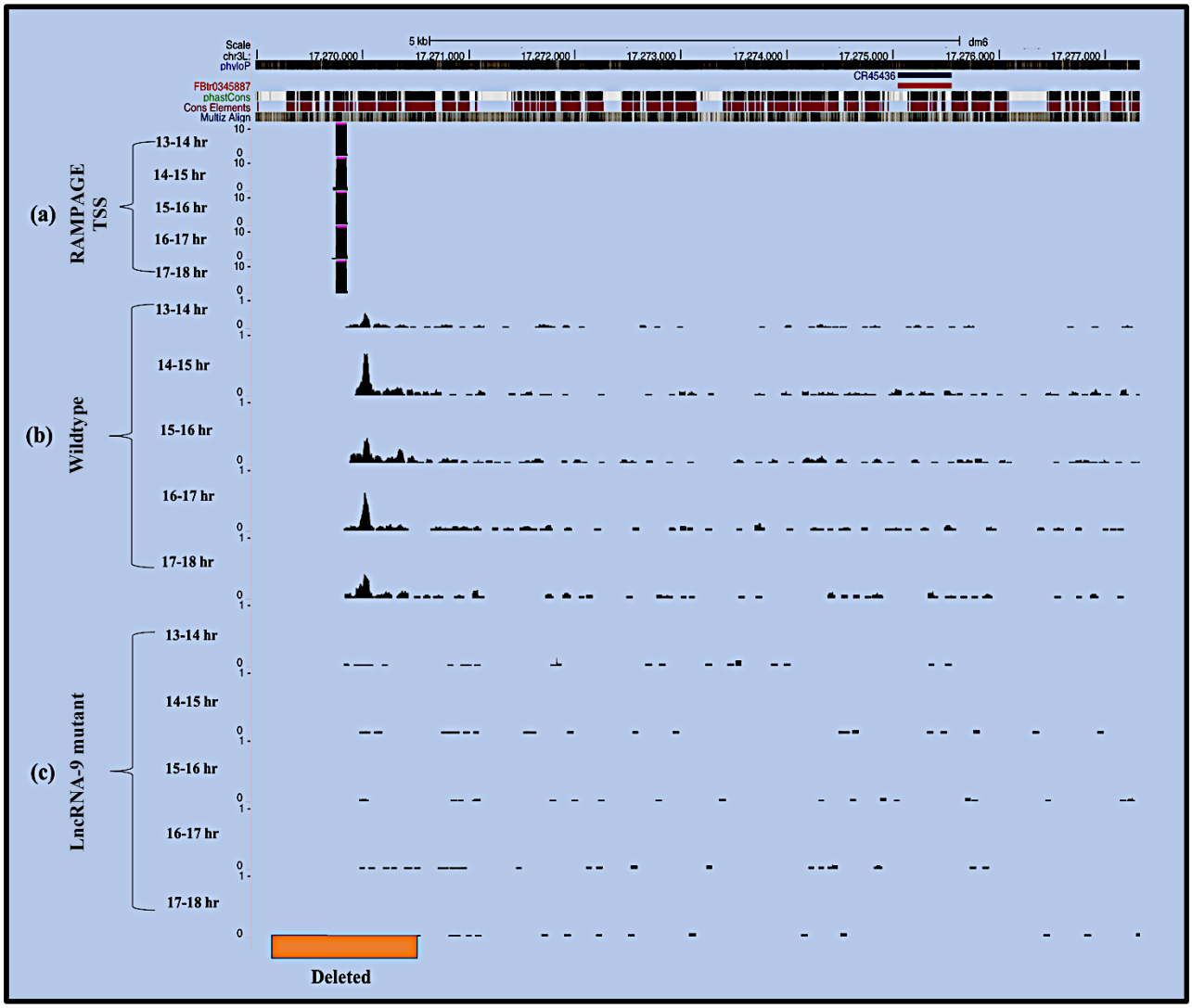
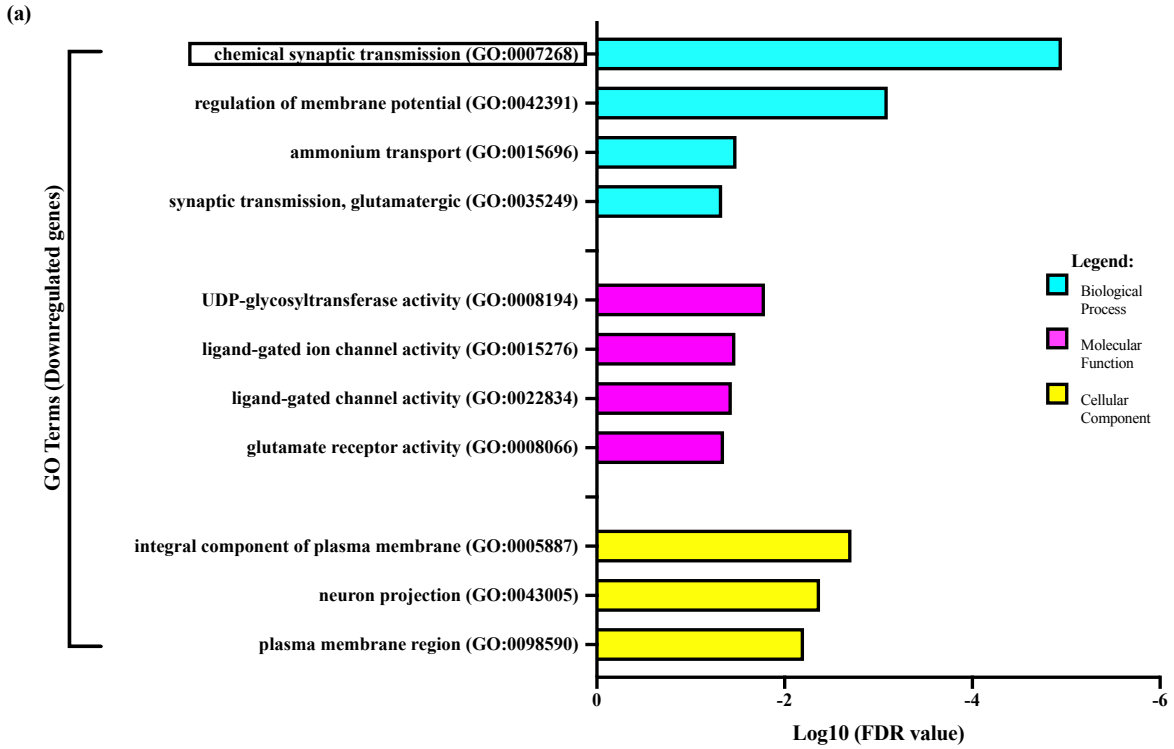


Figure 27: Genomic locus of lncRNA-9 from RNA-sequencing data.
All of the RNA-seq libraries were uploaded onto the UCSC genome browser and a portion of the RNA-seq libraries were shown here (chr3L: 17,269,000 – chr3L:17,277,300). The tracks for embryos collected from 13 hr–18 hr embryos were shown as it was the period of development where the lncRNA increasing faster before it reaches its highest expression at T16-T17. (a) As the lncRNA transcript was not annotated on FLYBASE or ENSEMBL, I made use of the RAMPAGE track to locate for the TSS of the lncRNA transcript (chr3L:17,269741 – chr3L:17,269766). (b) RNA-seq libraries generated with wildtype Oregon R flies. (c) RNA-seq libraries of homozygous null mutants. The results confirmed the KO of the lncRNA in the homozygous null mutant embryos. Orange bar indicates deleted region that was replaced with a 3xP3-RFP cassette.


(a)
GO Terms (Downregulated genes)
chemical synaptic transmission (GO:0007268)
regulation of membrane potential (GO:0042391)
ammonium transport (GO:0015696)
synaptic transmission, glutamatergic (GO:0035249)
UDP-glycosyltransferase activity (GO:0008194)
ligand-gated ion channel activity (GO:0015276)
ligand-gated channel activity (GO:0022834)
glutamate receptor activity (GO:0008066)
integral component of plasma membrane (GO:0005887)
neuron projection (GO:0043005)
plasma membrane region (GO:0098590)
Legend:
Biological Process
Molecular Function
Cellular Component
0
-2
-4
-6
Log10 (FDR value)

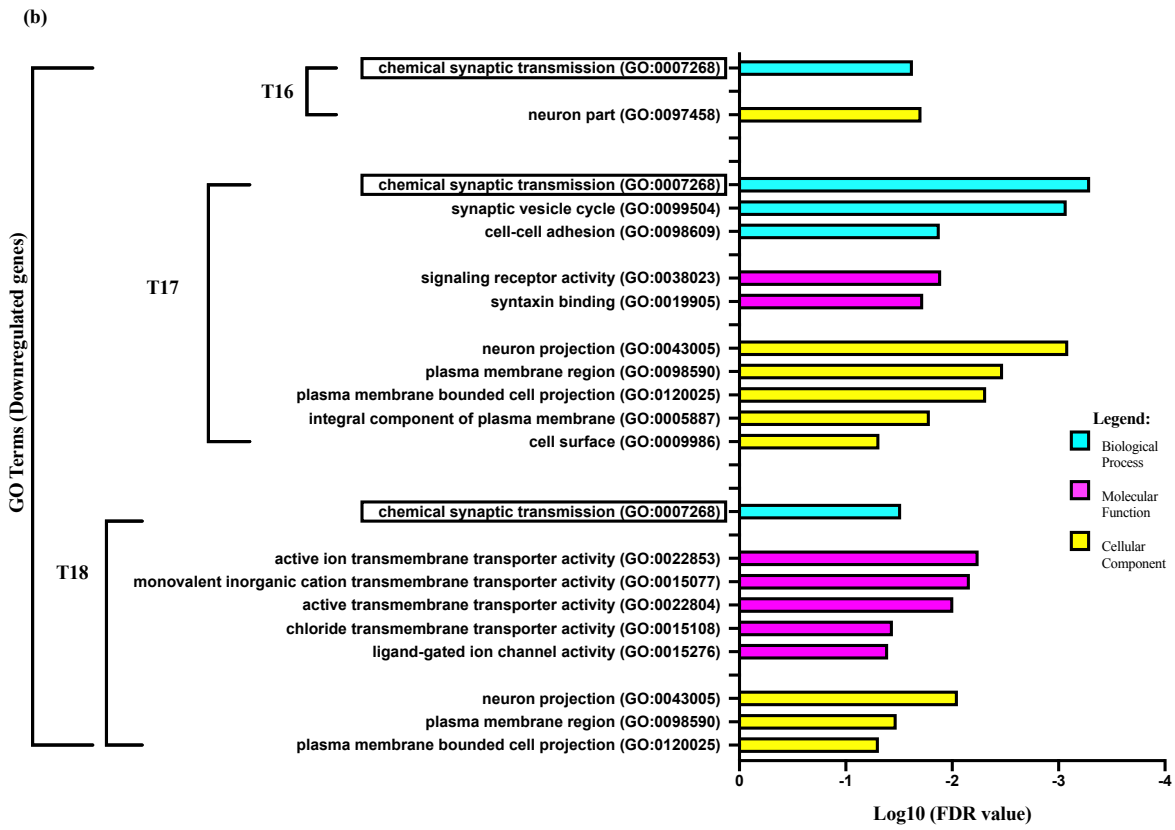

(b)
GO Terms (Downregulated genes)
T16
chemical synaptic transmission (GO:0007268)
neuron part (GO:0097458)
T17
chemical synaptic transmission (GO:0007268)
synaptic vesicle cycle (GO:0099504)
cell-cell adhesion (GO:0098609)
signaling receptor activity (GO:0038023)
syntaxin binding (GO:0019905)
neuron projection (GO:0043005)
plasma membrane region (GO:0098590)
plasma membrane bounded cell projection (GO:0120025)
integral component of plasma membrane (GO:0005887)
cell surface (GO:0009986)
T18
chemical synaptic transmission (GO:0007268)
active ion transmembrane transporter activity (GO:0022853)
monovalent inorganic cation transmembrane transporter activity (GO:0015077)
active transmembrane transporter activity (GO:0022804)
chloride transmembrane transporter activity (GO:0015108)
ligand-gated ion channel activity (GO:0015276)
neuron projection (GO:0043005)
plasma membrane region (GO:0098590)
plasma membrane bounded cell projection (GO:0120025)
Legend:
Biological Process
Molecular Function
Cellular Component
0
-1
-2
-3
-4
Log10 (FDR value)

Figure 28: *Drosophila* lncRNA-9 knockout reveals an impact on neuronal genes.
Depletion of lncRNA-9 results in downregulation of genes with important roles at the synapses. The plots are the results of functional annotation analysis of the top 200 downregulated genes based on their padj values. (a) GO analysis of genes that are downregulated at T10 our initial timepoint when the lncRNA is still lowly expressed. (b) GO analysis of genes at T16, T17 and T18. The lncRNA transcript peaks at T16 and decreases in expression levels after T18. Each GO domain has been coloured coded with Biological processes in cyan, molecular function in magenta and cellular components in yellow. In all of the timepoints, 'chemical synaptic transmission' is a common biological process with the lowest FDR. The enriched genes were also found to be localised with GO terms for neuronal activity (neuron projection, neuron part, ligand-gated ion activity).

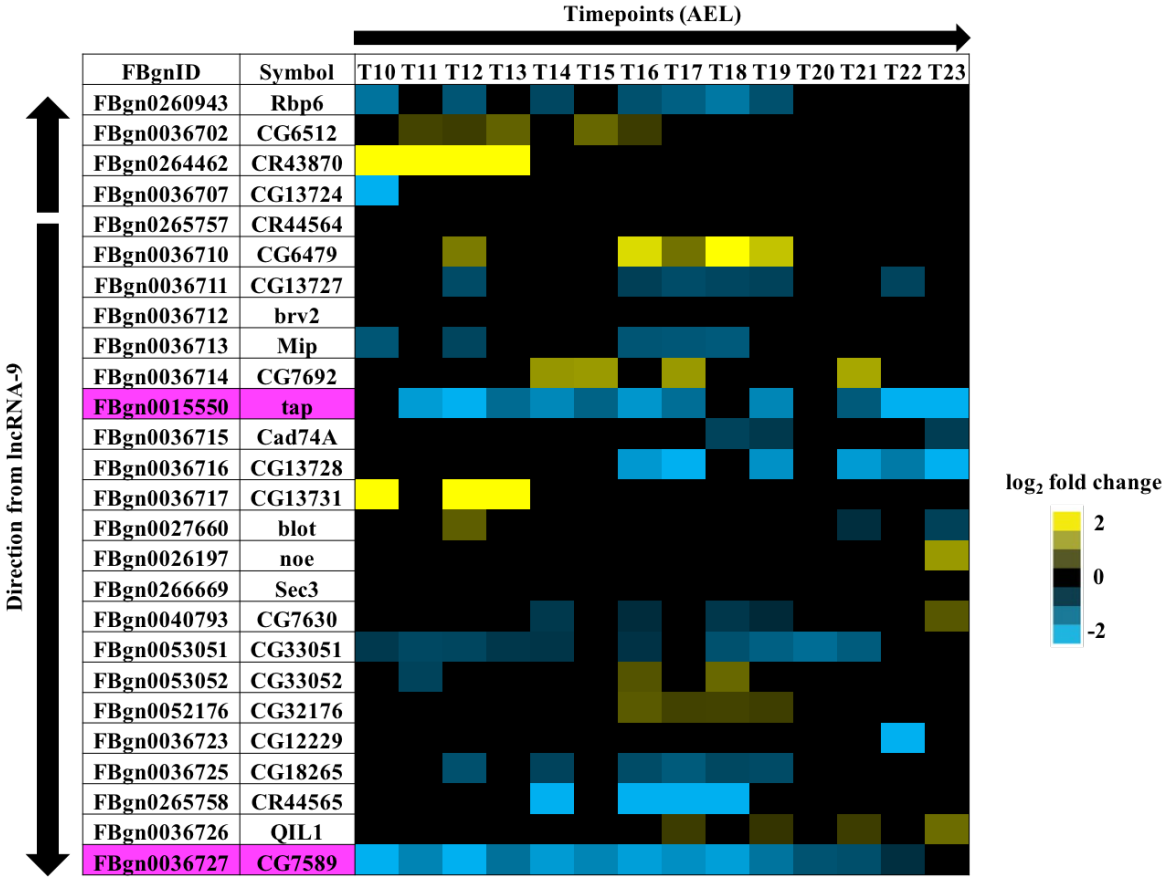


Figure 29: Differential gene expression profile of neighbouring genes from lncRNA-9.
RNA-seq heatmap of genes that are 200 kb from the genomic loci of the lncRNA transcript. The arrows represent the genomic location of the genes from lncRNA-9. The arrow pointing upwards indicates upstream whereas the arrow that is pointing downwards indicates downstream from the lncRNA. Genes that are closer to the arrowhead are the furthest away from the lncRNA locus. The heatmap has been plotted across indicated time points (14 time points, n = 2 biological replicates). The fold changes for each gene between lncRNA mutant and wildtype is represented in logarithmic (log_2) scale as shown. FBgn0015550 (*tap*) and FBgn0036727 (CG7589, *SecCl*) are genes (highlighted in magenta) involved in neuronal functions and found to be downregulated in the mutants in at least five timepoints consecutively (>1 fold, padj<0.05).

Table 17: modENCODE tissue expression data for lncRNA-3

#Library Collection ID	Category Descriptive Name	RPKM Expression Value
FBlc0000229	mE_mRNA_L3_Wand_imag_disc	2
FBlc0000225	mE_mRNA_L3_CNS	2
FBlc0000224	mE_mRNA_P8_CNS	1
FBlc0000210	mE_mRNA_A_VirF_1d_head	0
FBlc0000211	mE_mRNA_A_VirF_4d_head	0
FBlc0000231	mE_mRNA_A_VirF_20d_head	0
FBlc0000207	mE_mRNA_A_MateF_1d_head	1
FBlc0000213	mE_mRNA_A_MateF_4d_head	1
FBlc0000212	mE_mRNA_A_MateF_20d_head	0
FBlc0000209	mE_mRNA_A_MateM_1d_head	1
FBlc0000216	mE_mRNA_A_MateM_4d_head	1
FBlc0000214	mE_mRNA_A_MateM_20d_head	1
FBlc0000230	mE_mRNA_L3_Wand_saliv	0
FBlc0000234	mE_mRNA_WPP_saliv	0
FBlc0000227	mE_mRNA_L3_Wand_dig_sys	0
FBlc0000219	mE_mRNA_A_1d_dig_sys	0
FBlc0000223	mE_mRNA_A_4d_dig_sys	0
FBlc0000221	mE_mRNA_A_20d_dig_sys	0
FBlc0000228	mE_mRNA_L3_Wand_fat	0
FBlc0000233	mE_mRNA_WPP_fat	1
FBlc0000235	mE_mRNA_P8_fat	0
FBlc0000226	mE_mRNA_L3_Wand_carcass	1
FBlc0000218	mE_mRNA_A_1d_carcass	1
FBlc0000222	mE_mRNA_A_4d_carcass	1
FBlc0000220	mE_mRNA_A_20d_carcass	1
FBlc0000232	mE_mRNA_A_VirF_4d_ovary	0
FBlc0000208	mE_mRNA_A_MateF_4d_ovary	0
FBlc0000217	mE_mRNA_A_MateM_4d_testis	0
FBlc0000215	mE_mRNA_A_MateM_4d_acc_gland	0

Table 18: GO analysis for the top 200 DE genes with the lowest padj values (T10 – T23)

Time point	Biological Process	Molecular Function	Cellular Component
T10	• chemical synaptic transmission (GO:0007268) • ammonium transport (GO:0015696) • synaptic transmission, glutamatergic (GO:0035249)	*n.s.e	• integral component of plasma membrane (GO:0005887)
T11	• double-strand break repair via NHEJ (GO:0006303) • telomere maintenance (GO:0000723)	• telomeric DNA binding (GO:0042162)	• nuclear chromosome, telomeric region (GO:0000784)
T12	*n.s.e	• telomeric DNA binding (GO:0042162)	• nuclear chromosome, telomeric region (GO:0000784)
T13	*n.s.e	*n.s.e	*n.s.e
T14	*n.s.e	*n.s.e	*n.s.e
T15	*n.s.e	*n.s.e	*n.s.e
T16	*n.s.e	• telomeric DNA binding (GO:0042162)	• nuclear chromosome, telomeric region (GO:0000784) • nuclear chromosome part (GO:0044454)
T17	• synaptic vesicle cycle (GO:0099504) • chemical synaptic transmission (GO:0007268)	• syntaxin binding (GO:0019905)	• neuron projection (GO:0043005)
T18	• double-strand break repair (GO:0006302)	*n.s.e	• replisome (GO:0030894) • nuclear chromosome, telomeric region (GO:0000784) • DNA helicase complex (GO:0033202) • CMG complex (GO:0071162) • nuclear replication fork (GO:0043596) • origin recognition complex (GO:0000808) • transferase complex, transferring phosphorus-containing groups (GO:0061695)
T19	*n.s.e	telomeric DNA binding (GO:0042162)	• Pwp2p-containing subcomplex of 90S preribosome (GO:0034388) • nuclear chromosome, telomeric region (GO:0000784)
T20	*n.s.e	*n.s.e	*n.s.e
T21	*n.s.e	*n.s.e	*n.s.e
T22	*n.s.e	*n.s.e	*n.s.e
T23	*n.s.e	*n.s.e	• extracellular space (GO:0005615)

*n.s.e: no significant enrichment

Table 19.1: GO analysis for the top 200 downregulated genes with the lowest padj values (T10 – T18)

Time point	Biological Process	Molecular Function	Cellular Component
T10	• ammonium transport (GO:0015696) • synaptic transmission, glutamatergic (GO:0035249) • regulation of membrane potential (GO:0042391) • chemical synaptic transmission (GO:0007268)	• glutamate receptor activity (GO:0008066) • ligand-gated channel activity (GO:0022834) • ligand-gated ion channel activity (GO:0015276) • UDP-glycosyltransferase activity (GO:0008194)	• plasma membrane region (GO:0098590) • integral component of plasma membrane (GO:0005887) • neuron projection (GO:0043005)
T11	*n.s.e	• UDP-glycosyltransferase activity (GO:0008194)	*n.s.e
T12	*n.s.e	• UDP-glycosyltransferase activity (GO:0008194)	• intrinsic component of plasma membrane (GO:0031226)
T13	• fatty acid catabolic process (GO:0009062) • carboxylic acid catabolic process (GO:0046395) • cellular lipid catabolic process (GO:0044242)	• carboxylic acid binding (GO:0031406)	*n.s.e
T14	*n.s.e	*n.s.e	*n.s.e
T15	• carboxylic acid catabolic process (GO:0046395)	*n.s.e	*n.s.e
T16	• chemical synaptic transmission (GO:0007268)	*n.s.e	• neuron part (GO:0097458)
T17	• cell-cell adhesion (GO:0098609) • synaptic vesicle cycle (GO:0099504) • chemical synaptic transmission (GO:0007268)	• syntaxin binding (GO:0019905) • signalling receptor activity (GO:0038023)	• cell surface (GO:0009986) • integral component of plasma membrane (GO:0005887) • plasma membrane bounded cell projection (GO:0120025) • plasma membrane region (GO:0098590) • neuron projection (GO:0043005)
T18	• chemical synaptic transmission (GO:0007268)	• ligand-gated ion channel activity (GO:0015276) • chloride transmembrane transporter activity (GO:0015108) • active transmembrane transporter activity (GO:0022804) • monovalent inorganic cation transmembrane transporter activity (GO:0015077) • active ion transmembrane transporter activity (GO:0022853)	• plasma membrane bounded cell projection (GO:0120025) • plasma membrane region (GO:0098590) • neuron projection (GO:0043005)

*n.s.e: no significant enrichment

Table 19.2: GO analysis for the top 200 downregulated genes with the lowest padj values (T19 – T23)

Time point	Biological Process	Molecular Function	Cellular Component
T19	• ribosome biogenesis (GO:0042254) • ncRNA processing (GO:0034470) • rRNA processing (GO:0006364) • carboxylic acid catabolic process (GO:0046395) • ribosome biogenesis (GO:0042254)	• snoRNA binding (GO:0030515)	• small-subunit processome (GO:0032040) • Pwp2p-containing subcomplex of 90S preribosome (GO:0034388) • t-UTP complex (GO:0034455)
T20	• maturation of LSU-rRNA (GO:0000470) • maturation of 5.8S rRNA (GO:0000460)	• snoRNA binding (GO:0030515)	• Pwp2p-containing subcomplex of 90S preribosome (GO:0034388) • small nucleolar ribonucleoprotein complex (GO:0005732) • preribosome, large subunit precursor (GO:0030687) • small-subunit processome (GO:0032040) • t-UTP complex (GO:0034455)
T21	• rRNA processing (GO:0006364)	• snoRNA binding (GO:0030515)	• Pwp2p-containing subcomplex of 90S preribosome (GO:0034388) • small-subunit processome (GO:0032040) • t-UTP complex (GO:0034455)
T22	*n.s.e	*n.s.e	• extracellular matrix (GO:0031012) • protein-containing complex (GO:0032991) • extracellular space (GO:0005615) • intracellular organelle part (GO:0044446)
T23	*n.s.e	*n.s.e	• extracellular matrix (GO:0031012) • extracellular space (GO:0005615)

*n.s.e: no significant enrichment

Table 20.1: GO analysis for the top 200 upregulated genes with the lowest padj values (T10 – T16)

Time point	Biological Process	Molecular Function	Cellular Component
T10	*n.s.e	*n.s.e	*n.s.e
T11	• non-recombinational repair (GO:0000726) • double-strand break repair via nonhomologous end joining (GO:0006303) • telomere maintenance (GO:0000723) • telomere organization (GO:0032200)	• telomeric DNA binding (GO:0042162)	• site of double-strand break (GO:0035861) • nuclear chromosome, telomeric region (GO:0000784)
T12	• L-alpha-amino acid transmembrane transport (GO:1902475) • double-strand break repair via nonhomologous end joining (GO:0006303) • double-strand break repair (GO:0006302) • DNA duplex unwinding (GO:0032508)	• catalytic activity, acting on DNA (GO:0140097) • oxidoreductase activity (GO:0016491) • telomeric DNA binding (GO:0042162)	• nuclear chromosome, telomeric region (GO:0000784)
T13	*n.s.e	*n.s.e	*n.s.e
T14	• response to ionizing radiation (GO:0010212) • double-strand break repair via NHEJ (GO:0006303) • chromosome organization involved in meiotic cell cycle (GO:0070192) • telomere maintenance (GO:0000723)	• telomeric DNA binding (GO:0042162)	• condensed nuclear chromosome (GO:0000794) • nuclear chromosome part (GO:0044454) • nuclear chromosome, telomeric region (GO:0000784)
T15	*n.s.e	*n.s.e	*n.s.e
T16	• RNA metabolic process (GO:0016070) • DNA duplex unwinding (GO:0032508) • cellular response to abiotic stimulus (GO:0071214) • positive regulation of mitotic cell cycle (GO:0045931) • protein-DNA complex assembly (GO:0065004) • response to ionizing radiation (GO:0010212) • translesion synthesis (GO:0019985) • double-strand break repair via HR (GO:0000724) • telomere maintenance (GO:0000723) • DNA-dependent DNA replication (GO:0006261)	• chromatin binding (GO:0003682) • endodeoxyribonuclease activity (GO:0004520) • DNA-dependent ATPase activity (GO:0008094) • exonuclease activity (GO:0004527) • DNA-directed DNA polymerase activity (GO:0003887) • telomeric DNA binding (GO:0042162) • single-stranded DNA binding (GO:0003697)	• transferase complex, transferring phosphorus-containing groups (GO:0061695) • GINS complex (GO:0000811) • nuclear chromosome, telomeric region (GO:0000784) • origin recognition complex (GO:0000808) • nuclear chromatin (GO:0000790) • condensed nuclear chromosome (GO:0000794) • site of double-strand break (GO:0035861) • DNA helicase complex (GO:0033202) • replication fork (GO:0005657)

*n.s.e: no significant enrichment

Table 20.2: GO analysis for the top 200 upregulated genes with the lowest padj values (T17 – T18)

Time point	Biological Process	Molecular Function	Cellular Component
T17	*n.s.e	*n.s.e	• transferase complex, transferring phosphorus-containing groups (GO:0061695) • nuclear lumen (GO:0031981) • intracellular organelle lumen (GO:0070013)
T18	• crossover junction endodeoxyribonuclease activity (GO:0008821) • 3'-5'-exodeoxyribonuclease activity (GO:0008296) • nucleoside monophosphate kinase activity (GO:0050145) • damaged DNA binding (GO:0003684) • nucleosome binding (GO:0031491) • DNA-directed DNA polymerase activity (GO:0003887) • DNA-dependent ATPase activity (GO:0008094) • telomeric DNA binding (GO:0042162) • DNA replication origin binding (GO:0003688) • single-stranded DNA binding (GO:0003697)	• pre-replicative complex assembly involved in nuclear cell cycle DNA replication (GO:0006267) • mitotic DNA replication (GO:1902969) • mitotic recombination (GO:0006312) • regulation of mitotic cell cycle phase transition (GO:1901990) • protein K11-linked ubiquitination (GO:0070979) • double-strand break repair via break-induced replication (GO:0000727) • positive regulation of mitotic cell cycle (GO:0045931) • DNA strand elongation involved in DNA replication (GO:0006271) • DNA replication initiation (GO:0006270) • DNA replication (GO:0006260)	• nuclear chromatin (GO:0000790) • U5 snRNP (GO:0005682) • condensed nuclear chromosome (GO:0000794) • replisome (GO:0030894) • DNA helicase complex (GO:0033202) • U4 snRNP (GO:0005687) • nuclear pre-replicative complex (GO:0005656) • GINS complex (GO:0000811) • CMG complex (GO:0071162) • MCM core complex (GO:0097373)

*n.s.e: no significant enrichment

Table 20.3: GO analysis for the top 200 upregulated genes with the lowest padj values (T19 – T23)

Time point	Biological Process	Molecular Function	Cellular Component
T19	• response to ionizing radiation (GO:0010212) • meiotic nuclear division (GO:0140013) • mitotic sister chromatid segregation (GO:0000070) • cellular response to abiotic stimulus (GO:0071214) • nucleotide-excision repair (GO:0006289) • telomere maintenance (GO:0000723) • mitotic recombination (GO:0006312) • DNA-dependent DNA replication (GO:0006261) • double-strand break repair via homologous recombination (GO:0000724) • DNA biosynthetic process (GO:0071897)	• nucleoside monophosphate kinase activity (GO:0050145) • DNA-directed DNA polymerase activity (GO:0003887) • 3'-5' exonuclease activity (GO:0008408) • endodeoxyribonuclease activity (GO:0004520) • telomeric DNA binding (GO:0042162) • damaged DNA binding (GO:0003684) • single-stranded DNA binding (GO:0003697)	• U4/U6 x U5 tri-snRNP complex (GO:0046540) • DNA helicase complex (GO:0033202) • MCM core complex (GO:0097373) • nuclear chromosome, telomeric region (GO:0000784) • U4 snRNP (GO:0005687) • replisome (GO:0030894) • GINS complex (GO:0000811) • CMG complex (GO:0071162) • transferase complex, transferring phosphorus-containing groups (GO:0061695) • replication fork (GO:0005657)
T20	*n.s.e	*n.s.e	*n.s.e
T21	*n.s.e	*n.s.e	*n.s.e
T22	*n.s.e	*n.s.e	*n.s.e
T23	• small molecule metabolic process (GO:0044281)	*n.s.e	*n.s.e

*n.s.e: no significant enrichment

4.3 Discussion

Given the rigorous control that is required to rapidly transform an embryo from an unrelated lineage into larvae with fully developed organs, well-coordinated processes consisting of a myriad of factors are needed to establish the new cell identities. In this study, following the identification of two candidate lncRNAs that presented homozygous lethal phenotypes, I presented a more detailed characterisation of the molecular consequences for the loss of one of them. To date, thousands of lncRNA transcripts have been identified but few have been characterised. Despite the increasing number of transcriptomic studies generated over the years, specific mechanistic function of lncRNAs and their mode of action remained unidentified. To address these questions, we made use of mutants of lncRNA candidates that were selected based on RAMPAGE data (Batut *et al.,* 2017) in chapter 3. *Drosophila* has been used widely over last century in genetic research and characterisation of the various stages of development. Embryogenesis, specifically, can be considered to be one of the most well-studied ones. To study the function of these lncRNAs in vivo, I made use of CRISPR-Cas9 technology to remove the TSS region and replaced it with a fluorescence marker. The two lncRNA candidates, lncRNA-9 and lncRNA-3*,* were found to be embryonic lethal and pupal lethal, respectively.

Further to generating these flies, I have taken initial steps towards characterising the tissue expression of these two lncRNAs using publicly available data. Our initial information was limited to their genomic loci and their expression in the adult head. The brain is a fascinating organ that have been extensively studied with one of the greatest numbers of unique cell-type specific lncRNAs (Qureshi *et al.,* 2010; Andersen *et al.,* 2018). To investigate the localisation of these lncRNA candidates, we made use of the modENCODE database that included ChIP-Seq and RNA sequencing data from multiple tissues and developmental stages of *Drosophila.* Besides the adult head, the lncRNA-3 transcript was found to be highly expressed in the neurons and neuroblast of the central nervous system of L3 larvae. Compared to lncRNA-3*,* the lncRNA-9 transcript was found to have low-moderate expression in the adult head and neurons of the central nervous system. With the increasing number of lncRNA transcripts heir expression and specificity. Chromatin marks that surround the promoter region of each individual lncRNA have been characterised. The authors had made use of a computational approach which utilises RNA-seq and ChIP-seq datasets to improve the annotation of curated lncRNAs (Chen *et al.,* 2016). While it is reported that an absence of chromatin signatures related to active transcription were seen in more than half of the curated lncRNAs, further

investigations will be needed to understand the impact that these regulatory elements have in defining one type of lncRNA from another which would affect its biological functions. In addition, little is known about the impact that the chromatin marks have on the expression of the lncRNA genomic loci and more specific genome-wide investigation is needed to understand their operation and the selective constraints that have acted on the lncRNA.

During this study, another important question that I addressed was whether lncRNA-3 and lncRNA-9 exert their effect in a cis or trans. A lncRNA gene whose primary function is exerted in trans would be rescued by the expression of the lncRNA on a separate chromosome whereas a cis-acting lncRNA will not. I therefore generated flies that expressed BAC rescue clones for each of the lncRNA from other chromosomes than the lncRNA locus. When expressed from a BAC transgene, lncRNA-3 and lncRNA-9 did not rescue the lethal phenotype caused by lncRNA deletion. These results suggest that lncRNA-3 and lncRNA-9 function in cis rather than in a trans-acting manner.

In this chapter, a further investigation of the molecular phenotype of lncRNA-9 was carried out. To have a better understanding of the subcellular localisation of the lncRNA transcript, in situ RNA fluorescence hybridization assays were used to study the localisation of lncRNA-9 (Raj *et al.,* 2008; Crosetto *et al.,* 2015). RNA fluorescence in situ hybridization (RNA FISH) allowed me to directly assess the expression and subcellular localization of lncRNA-9. RNA-FISH experiments from staged embryos using oligo probe sets from RNAscope and single-molecule hybridization chain reaction (HCR) were used to determine the localisation of lncRNA-9 within the embryo (Wang *et al.,* 2012; Choi *et al.,* 2016). Our results indicated that lncRNA-9 is localised in the nucleus and co-expressed with specific neuronal cells in the ventral nerve cord. This finding confirmed the information obtained from modENCODE that the lncRNA was detected in the central nervous system and in neurons. The advantage of using both of these RNA FISH methods is their ability to amplify the signal to produce a stronger fluorescence and higher signal-to-noise ratios. In addition, both methods gave us better confidence of the results as similar transcript localisation was observed. I have previously used single-molecule RNA FISH (smFISH) oligo probes obtained from Stellaris but due to the lack of transcript structure information, we were only able to generate less than 25 oligonucleotides for each lncRNA candidate (minimum requirement was 25 per gene). Low fluorescence smFISH signals combined with high background autofluorescence in *Drosophila* embryos made it challenging to obtain any localisation information with this oligo probe set (results not shown). I have settled for HCR probes from molecular technologies as

the method had published protocols that are suitable for whole-mount embryos and it allowed us to perform multiplexed experiments with antibodies. Given the success with the HCR method for lncRNA-9, we also aim to decipher the localisation of lncRNA-3 in the developing embryo and adult CNS in the future once we have validated the phenotype of the lncRNA. While RNA FISH is a very powerful technique for the analysis of lncRNA localisation, extensive troubleshooting and validation strategies such as the use of alternative techniques or probing with extra of oligo sets is crucial to ensuring the validity of RNA FISH signals.

To understand the downstream genes that are affected by lncRNA-9 loss, RNA-sequencing libraries were generated using wildtype and homozygous null mutant embryos collected at hourly time points that corresponded to the expression profile of the lncRNA. Our preliminary DESeq analysis revealed a list of genes whose expression were significantly changed upon the loss of the lncRNA transcript at each timepoint. Much focus was placed at examining a small subset of genes. To enrich for gene clusters within the list of genes, two approaches were used. I first performed a GO analysis using gene lists that were filtered for the top 200 genes with the largest changes in padj values. The second approach took into account the log fold changes of genes (downregulated or upregulated) before ranking them. GO enrichment analysis revealed clusters of genes related to *Drosophila* nervous system processes such as ion transmembrane channel activities, neuron projection and chemical synapse transmission. Furthermore, one of the candidates from the pathway analysis, *SecCl*, is located approximately 200 kbp downstream from the lncRNA transcript and its expression is downregulated upon the loss of lncRNA-9. Furthermore, several glycine receptors were also downregulated in the same class. Due to the large distance of *SecCl* from the transcriptional start site of lncRNA-9, further validation will be needed to examine if this is due to secondary effects. Previous studies made use of neighbouring mRNA− lncRNA pairing to identify lncRNAs that act in cis. Most of these approaches were performed computationally due to the lack of systematic ways of performing this task (Cabili *et al.,* 2011; Sigova *et al.,* 2013). It was also shown that actual physical interaction between two genes is not limited by the genomic distance (Khyzha *et al.,* 2019) and other factors have to be taken into consideration.

In this study, I made use of the information derived from the GO consortium database to search for functional enrichment. Each of the principal GO domains consists of a hierarchy of terms that can be of broad functions down to specific levels. While there are many advantages for using GO term analyses for functional enrichment, there remain some drawbacks of this approach. One of the major drawbacks is the quality of the databases and

whether it is manually curated or purely based on predictions that could skew the results. The use of high-quality ontologies is essential and it is noteworthy that it is not always accurate to assume that the genes in a group are associated with each other. The second issues with this approach is that I have only selected for the genes that have high padj values and during this process of selection, many genes whose expression is moderately changed will be omitted. While it is a common approach to look for the most differentially expressed genes or those with the most significant padj values, most of the pathways in the GO analysis are made up of several components rather than a single factor. It is thus a common problem that during the analysis, no significant enrichment was detected.

While the results presented here show that lncRNA-9 is expressed in some neurons in the CNS and suggest that is likely able to influence synaptic transmission, we were not yet able to establish how the lncRNA drives the observed downstream changes in gene expression. One potential candidate might be Tap, which is very close to the lncRNA and being a transcription factor could be responsible for changes in many downstream targets. Likewise, we cannot formally confirm whether these signatures observed, especially for genes surrounding the lncRNA are directly caused by the lncRNA absence (i.e. they are primary targets of lncRNA-9) or a by-product of arrested development and defects in neuronal cells in the lncRNA mutant. Further validation work will be required to determine how this ncRNA mediates these changes and which changes are ultimately responsible for the lethality. We are in the process of testing various antibodies for neurons and synapses of the nervous system to perform combined RNA-FISH and IF experiments to look at their localisation in the embryo alongside lncRNA-9. Additionally, to study the biochemical partners of the lncRNA that mediate this phenotype, methods such as RNA Antisense Purification with Mass Spectrometry (RAP-MS) would reveal lncRNA interacting partners involved in regulating different process (McHugh and Guttman, 2018). Similarly, methods such as RNA-DAMID and CHART-seq have been used to identify genomic binding sites of *Drosophila* roX2 lncRNAs and these protocols would be helpful for us given their proven protocols for identifying the binding sites of a lncRNA in *Drosophila* (Simon et al., 2011; Cheetham and Brand, 2018). To perform these experiments, we would need to determine the full-length transcript of lncRNA-9 with methods such as 5’ and 3’ RACE as well as Northern Blot analysis. Noteworthy, we have not been successful in using computational de novo assembly methods such as Trinity to determine the full-length transcript.

Another open question that will remain to be addressed is how the changes observed ultimately lead to the lethality observed in the embryos. Even in this respect, Tap might be a given that the phenotype for flies lacking Tap is closely similar that observed in lncRNA-9 mutants. Even though there are a few escapers to early larval stages, the majority of the Tap mutants were reported as embryonic lethal (Yuan *et al.,* 2016). To address this question a rescue experiment could be attempted to check if Tap overexpression is able to rescue the phenotype produced by lncRNA-9 absence.

Overall, using various experiments and analyses, I have attempted to functionally characterise the candidate lncRNA-9. Due to the challenging nature of research on lncRNAs, further experimental efforts will be necessary to gain a full understanding of how these ncRNAs act. Within the *Drosophila* CNS, different cells and genes work in synchrony in this complex system. Many lncRNA transcripts have been identified and suggested to function in the nervous system. Similar to lncRNA-9, the other candidate, lncRNA-3, was found to be highly expressed in the CNS and it would be interesting to understand the functional relevance of the lncRNA during *Drosophila* nervous system development with future experiments.

Supplementary Table 1.1: Top 100 most DE genes with the lowest padj values (T10)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.48E-72	-5.415
2	CG5687	9.66E-50	4.917
3	w	5.49E-48	-4.729
4	mthl8	1.06E-47	-8.077
5	Grik	1.13E-42	-6.430
6	lncRNA:CR43459	2.70E-33	-2.815
7	Drsl5	2.46E-31	10.500
8	ZnT77C	3.32E-30	-2.605
9	CG33337	4.52E-30	-8.391
10	CG11319	1.14E-27	-2.771
11	AstC-R2	1.00E-26	-3.354
12	CG14990	5.03E-25	-3.361
13	CG31098	1.11E-24	2.392
14	asRNA:CR45835	1.89E-24	-5.508
15	SecCl	1.89E-24	-2.422
16	CG9360	7.61E-24	-4.968
17	Cyp4p2	1.52E-23	-3.466
18	CR42490	1.88E-23	-7.155
19	CG31157	5.50E-22	-4.925
20	asRNA:CR45822	7.92E-22	-9.710
21	asRNA:CR44029	1.14E-21	-3.618
22	CG42822	2.92E-21	4.245
23	lncRNA:CR44466	2.59E-20	-3.699
24	dpr20	4.60E-20	-2.459
25	CG42365	1.30E-18	-3.244
26	CG11236	1.77E-17	-3.078
27	lncRNA:CR44754	1.89E-17	-3.561
28	CG31288	5.32E-17	3.671
29	Cyp12c1	5.98E-17	-3.330
30	mthl4	6.19E-17	-2.731
31	CG14301	6.19E-17	-2.345
32	CG32195	8.33E-17	-2.381
33	nAChRalpha7	1.86E-16	-2.410
34	mGluR	5.47E-16	-2.363
35	NLaz	1.36E-15	-3.099
36	CG14356	1.86E-15	-3.244
37	CG18547	2.17E-15	1.632
38	CG10445	2.20E-15	3.352
39	CG14204	2.30E-15	-4.167
40	CG31676	2.36E-15	-2.110
41	yellow-h	2.57E-15	-2.258
42	CG33468	3.50E-15	-3.941
43	nompC	4.03E-15	-1.919
44	St3	7.58E-15	-2.511
45	TyrR	1.03E-14	2.293
46	Scgalpha	1.17E-14	-2.292
47	CG31997	2.00E-14	-2.233
48	CG9568	2.51E-14	-1.949
49	Ect3	4.79E-14	-2.122
50	AOX1	4.99E-14	1.831

Rank	Gene Symbol	padj value	log2 fc
51	GILT2	6.37E-14	1.162
52	lncRNA:CR44536	6.70E-14	9.477
53	asRNA:CR45140	7.21E-14	-3.122
54	CG43338	1.04E-13	-1.946
55	Ugt35A1	1.23E-13	-1.607
56	CG42337	1.68E-13	-1.865
57	CG10407	1.89E-13	-1.964
58	lncRNA:CR46075	2.52E-13	-4.143
59	CG2064	2.55E-13	2.279
60	CR44383	4.78E-13	-5.397
61	asRNA:CR44960	5.21E-13	4.839
62	CG34155	5.61E-13	-2.170
63	lncRNA:CR44743	6.75E-13	-3.970
64	prom	6.75E-13	6.116
65	Cyp9c1	8.83E-13	-2.575
66	CR43105	8.83E-13	6.337
67	CG13033	1.12E-12	3.921
68	CG30059	1.13E-12	3.166
69	Ugt302K1	1.30E-12	-6.782
70	CG34354	1.30E-12	-2.050
71	lncRNA:CR45127	2.04E-12	-6.788
72	CG12493	3.38E-12	3.152
73	CG4752	3.68E-12	-1.601
74	Cib2	4.04E-12	-2.107
75	CG18278	4.13E-12	3.007
76	CG34384	5.83E-12	-1.677
77	CG43155	8.23E-12	-2.520
78	Est-6	8.31E-12	3.376
79	Cngl	1.11E-11	-1.772
80	CG3842	1.23E-11	-1.783
81	CG33128	1.53E-11	-5.242
82	Ugt303B3	2.35E-11	-2.360
83	CG16799	2.54E-11	-1.943
84	CG15515	3.98E-11	4.901
85	narya	4.45E-11	6.225
86	yellow-d2	8.42E-11	-4.515
87	lncRNA:CR46006	8.43E-11	-1.787
88	dsb	8.95E-11	-3.181
89	lncRNA:CR44467	1.00E-10	-2.540
90	Ptp52F	1.08E-10	2.283
91	CG1494	1.08E-10	-2.240
92	CG6044	1.18E-10	-1.750
93	tgy	1.59E-10	-2.073
94	CG1358	2.86E-10	2.062
95	mmd	3.86E-10	-1.654
96	Cyp4e3	4.24E-10	3.486
97	Had2	4.58E-10	-3.141
98	lncRNA:CR31781	5.08E-10	-1.757
99	List	5.36E-10	2.495
100	unc79	5.73E-10	-1.561

Supplementary Table 1.2: Top 100 most DE genes with the lowest padj values (T11)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.26E-70	-5.496
2	CG5687	7.85E-68	5.880
3	mthl8	6.82E-61	-6.800
4	w	9.73E-52	-4.932
5	CG18547	2.19E-41	2.660
6	Grik	1.13E-36	-5.499
7	CG31098	3.50E-31	2.715
8	CG33337	1.07E-30	-7.625
9	CG42822	3.14E-30	4.829
10	CG31288	7.26E-28	5.053
11	Drsl5	8.81E-28	9.519
12	lncRNA:CR44536	1.29E-27	8.828
13	CR42490	9.96E-26	-7.018
14	AOX1	2.30E-25	2.477
15	ZnT77C	1.02E-24	-2.352
16	asRNA:CR45835	3.37E-24	-4.658
17	CG2064	2.31E-23	3.021
18	CG1358	8.40E-23	3.211
19	asRNA:CR45822	1.22E-22	-9.914
20	CG11236	1.25E-22	-4.285
21	lncRNA:CR44754	4.89E-22	-3.700
22	Mocs1	5.78E-22	2.919
23	agt	1.66E-21	3.768
24	asRNA:CR44029	9.50E-21	-3.507
25	CG31157	5.07E-20	-4.540
26	lncRNA:CR43459	6.30E-20	-2.163
27	CG14204	7.41E-19	-5.443
28	CG3448	3.27E-18	2.509
29	sni	1.43E-17	1.552
30	frm	1.81E-17	2.653
31	CG42319	2.34E-17	2.770
32	rad50	2.77E-17	2.656
33	CG9360	3.87E-17	-4.242
34	Cyp4p2	1.91E-16	-2.805
35	asRNA:CR45140	2.11E-16	-3.485
36	CG10445	3.47E-16	3.415
37	CG14301	3.59E-16	-2.354
38	CG12766	8.83E-16	5.035
39	Nop17l	1.18E-15	-1.194
40	CG7054	2.01E-15	2.356
41	lncRNA:CR46075	2.17E-15	-4.615
42	CR43105	2.54E-15	7.038
43	Ect3	2.56E-15	-2.213
44	TwdlG	3.03E-15	3.168
45	CG31676	3.25E-15	-2.103
46	CG32694	3.67E-15	2.826
47	CG12224	5.71E-15	-2.072
48	CG10407	5.71E-15	2.895
49	CG14907	9.34E-15	2.034
50	bnk	3.74E-14	2.096

Rank	Gene Symbol	padj value	log2 fc
51	Ku80	3.88E-14	2.689
52	DNApol-zeta	4.19E-14	2.031
53	LManV	5.18E-14	3.654
54	Rev7	7.28E-14	2.430
55	ms(3)76Ba	7.35E-14	7.292
56	Argk	1.08E-13	-2.096
57	SP1029	1.71E-13	1.205
58	TyrR	1.77E-13	2.226
59	mthl4	1.85E-13	-2.373
60	Cyp12d1-d	3.89E-13	3.904
61	asRNA:CR45171	3.89E-13	4.756
62	CG17904	4.76E-13	1.113
63	zye	4.76E-13	2.981
64	Mco1	7.51E-13	2.717
65	Cpr62Bb	7.70E-13	3.001
66	drd	9.18E-13	2.507
67	CG31633	1.23E-12	1.339
68	Irbp	1.78E-12	2.497
69	CR40190	1.86E-12	-1.303
70	CG2909	2.01E-12	2.586
71	asRNA:CR44960	3.79E-12	4.823
72	lncRNA:CR44466	3.90E-12	-2.449
73	CG3520	4.25E-12	2.075
74	FoxP	4.58E-12	-2.238
75	CG15553	6.20E-12	-4.886
76	Cyp12d1-p	8.43E-12	3.879
77	CG14990	9.42E-12	-2.159
78	Tina-1	9.90E-12	-1.200
79	CG5704	1.22E-11	2.528
80	CG30427	1.23E-11	2.196
81	CG14695	1.47E-11	3.443
82	Blos1	1.72E-11	-0.961
83	smp-30	2.17E-11	-2.513
84	Spn77Bc	2.19E-11	-3.199
85	NijC	2.19E-11	-1.705
86	CG4847	2.46E-11	2.434
87	Cyp9b2	2.48E-11	-2.884
88	lncRNA:CR45242	2.48E-11	2.473
89	CG3165	2.56E-11	1.850
90	obe	2.82E-11	1.388
91	narya	2.82E-11	6.692
92	mthl3	3.78E-11	2.854
93	lncRNA:CR44743	4.47E-11	-3.456
94	CG14694	5.81E-11	-3.007
95	CG33281	5.83E-11	2.717
96	CG15098	9.43E-11	2.489
97	slif	1.10E-10	2.088
98	CG7607	1.13E-10	-2.259
99	Cyp6a13	1.42E-10	1.755
100	CG10638	1.79E-10	1.634

Supplementary Table 1.3: Top 100 most DE genes with the lowest padj values (T12)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.41E-58	-5.053
2	mthl8	1.95E-55	-6.770
3	w	1.24E-48	-4.809
4	CG5687	4.05E-41	4.426
5	asRNA:CR45822	2.21E-37	-8.646
6	Grik	4.04E-30	-5.009
7	lncRNA:CR43459	4.88E-25	-2.444
8	asRNA:CR44029	1.20E-24	-3.974
9	CG18547	9.63E-24	2.023
10	sni	2.31E-23	1.788
11	CG11319	3.54E-23	-2.310
12	CG10445	2.29E-22	4.210
13	CG10407	3.34E-22	-2.611
14	CG9360	4.76E-22	-4.846
15	CG18278	5.76E-21	4.132
16	CG42822	1.42E-20	3.666
17	Cpr56F	5.34E-20	3.928
18	asRNA:CR45140	1.48E-19	-4.254
19	asRNA:CR45835	2.70E-19	-4.427
20	Mco1	8.04E-19	2.104
21	CG2064	8.04E-19	2.740
22	CG31098	8.04E-19	3.890
23	CG11236	5.70E-18	-4.081
24	CG30059	3.97E-17	3.716
25	Fancl	8.93E-17	2.643
26	SecCl	9.61E-17	-2.003
27	AOX1	1.00E-16	2.024
28	ZnT77C	1.35E-16	-1.941
29	Picot	2.47E-16	2.825
30	CG31288	3.91E-16	3.622
31	lncRNA:CR44466	9.31E-16	-3.020
32	lncRNA:CR44536	1.15E-15	-2.246
33	Ect3	1.15E-15	8.399
34	mthl3	1.19E-15	3.651
35	CR43105	1.32E-15	5.917
36	CG33337	3.99E-15	-6.973
37	CG11584	4.88E-15	5.627
38	slif	5.18E-15	2.461
39	CG31676	7.35E-15	-2.059
40	rad50	1.03E-14	2.459
41	CG14204	1.15E-14	-4.704
42	aralar1	1.31E-14	1.781
43	Cpr67Fa1	1.47E-14	5.803
44	CG1358	2.19E-14	2.409
45	CG4730	2.28E-14	2.963
46	CG4842	2.32E-14	2.960
47	tut	3.75E-14	3.947
48	lncRNA:CR44743	4.78E-14	-4.355
49	CG3165	5.22E-14	2.070
50	CG14990	5.67E-14	-2.370

Rank	Gene Symbol	padj value	log2 fc
51	Tsp74F	5.96E-14	1.855
52	CG31157	7.55E-14	-3.764
53	NLaz	2.27E-13	-2.878
54	CG6293	3.53E-13	3.253
55	Scgalpha	4.96E-13	-2.155
56	narya	4.96E-13	7.707
57	Drsl5	5.01E-13	10.137
58	Had2	1.08E-12	-3.408
59	alpha-Est2	2.55E-12	-2.108
60	lncRNA:CR45242	2.86E-12	-3.278
61	zye	2.86E-12	2.936
62	lncRNA:CR44754	3.03E-12	-3.603
63	CG15820	4.13E-12	-2.302
64	FoxP	4.13E-12	6.035
65	Cyp12d1-d	4.60E-12	3.521
66	CG4267	4.69E-12	3.383
67	CG33288	5.19E-12	-2.131
68	AstC-R2	8.04E-12	-2.226
69	DNApol-zeta	9.04E-12	1.852
70	Cyp4p2	1.18E-11	-2.359
71	CG14301	1.18E-11	-2.284
72	CG31626	1.23E-11	3.383
73	CG13046	1.23E-11	4.622
74	CG12766	1.69E-11	3.716
75	Ku80	1.75E-11	2.436
76	CG8560	1.75E-11	-2.512
77	asRNA:CR45171	1.96E-11	3.960
78	NijC	2.30E-11	-1.757
79	Gk2	2.64E-11	1.485
80	Calx	2.64E-11	2.020
81	ms(3)76Ba	2.93E-11	8.744
82	Toll-9	3.00E-11	-2.969
83	CG8768	3.42E-11	-1.855
84	Spn88Ea	3.56E-11	1.449
85	CG4766	3.71E-11	-2.185
86	prom	3.79E-11	4.396
87	CG14882	3.88E-11	1.478
88	CG15649	4.05E-11	4.628
89	Argk	4.32E-11	-1.882
90	CG17292	6.97E-11	1.337
91	CG33128	8.08E-11	-4.607
92	Tb	8.51E-11	5.841
93	Irbp	8.68E-11	2.309
94	lncRNA:CR46006	9.10E-11	-1.788
95	CG3842	1.05E-10	-1.705
96	CG32694	1.15E-10	2.371
97	TwdlB	1.37E-10	5.091
98	Irk1	1.68E-10	1.222
99	CG42337	1.81E-10	-1.489
100	CG6180	2.74E-10	1.841

Supplementary Table 1.4: Top 100 most DE genes with the lowest padj values (T13)

Rank	Gene Symbol	padj value	log2 fc
1	CG5687	1.15E-50	4.729
2	mthl8	3.00E-45	-7.019
3	CG18547	1.86E-44	2.744
4	Hsc70-2	9.29E-41	-3.913
5	CG42822	3.46E-39	5.615
6	w	1.22E-35	-4.060
7	TwdlU	2.77E-32	-5.023
8	CG2064	3.39E-29	3.390
9	CG11236	9.30E-26	-4.210
10	lncRNA:CR44536	1.06E-24	9.243
11	Drsl5	2.64E-23	10.139
12	CG2233	3.10E-21	3.974
13	CG10407	8.84E-21	-2.584
14	lncRNA:CR43459	3.91E-19	-2.131
15	asRNA:CR44029	1.42E-18	-3.435
16	CG11319	3.30E-17	-1.894
17	CG31288	3.30E-17	3.175
18	CG12766	6.00E-17	4.156
19	Cyp6a18	6.05E-17	-3.002
20	CR43105	1.07E-16	5.807
21	ppk	1.11E-16	3.968
22	CG3165	3.35E-16	2.241
23	NLaz	7.10E-16	-3.141
24	CG31098	5.39E-15	1.838
25	mthl3	1.17E-14	3.270
26	Cyp6a13	1.19E-14	1.792
27	asRNA:CR44030	3.89E-14	3.241
28	CG33337	6.02E-14	-5.668
29	asRNA:CR45835	1.17E-13	-3.317
30	CG8369	1.57E-13	3.714
31	CG6163	1.67E-13	-3.402
32	Odc1	2.03E-13	3.518
33	asRNA:CR45140	3.79E-13	-3.624
34	AOX1	7.32E-13	1.756
35	CG3397	1.28E-12	2.707
36	ms(3)76Ba	2.07E-12	7.908
37	asRNA:CR45171	2.22E-12	4.212
38	CG31706	3.08E-12	5.484
39	SCAP	6.72E-12	1.806
40	CG15515	6.72E-12	5.110
41	Nop17l	7.36E-12	-1.054
42	CG18278	8.53E-12	3.139
43	Prp31	9.62E-12	1.479
44	CG17904	1.83E-11	1.060
45	CG13694	2.00E-11	2.871
46	CG4730	2.38E-11	2.634
47	CR43186	2.98E-11	4.486
48	asRNA:CR45822	3.75E-11	-9.048
49	Ku80	4.39E-11	2.411
50	CG33128	4.53E-11	-4.663
51	asRNA:CR44960	5.46E-11	4.637
52	CG9521	6.18E-11	6.192
53	sni	6.58E-11	1.200
54	CG2909	6.62E-11	2.459
55	Cpr76Bc	1.04E-10	-2.415
56	CG10799	1.07E-10	3.061
57	CG10638	1.60E-10	1.657
58	CG31676	1.63E-10	-1.689
59	narya	1.68E-10	6.528
60	Rev7	2.17E-10	2.061
61	ZnT35C	2.30E-10	1.675
62	CG9360	2.67E-10	-3.294
63	CG31157	3.24E-10	-3.338
64	alpha-Est2	4.62E-10	-1.749
65	Bin1	7.42E-10	1.685
66	CG30059	1.02E-09	2.865
67	lncRNA:CR44754	1.30E-09	-2.486
68	Had2	1.50E-09	-2.882
69	CG31606	1.71E-09	4.812
70	Bace	1.83E-09	3.289
71	CG7465	2.24E-09	5.482
72	lncRNA:CR45039	2.28E-09	3.693
73	CG3842	2.80E-09	-1.609
74	CG10960	3.41E-09	-1.086
75	Cpr11A	3.56E-09	2.389
76	prom	4.07E-09	3.574
77	Grik	4.56E-09	-2.613
78	LManV	6.15E-09	2.546
79	CR43697	7.79E-09	2.172
80	CG8541	8.11E-09	3.409
81	mthl4	8.94E-09	-1.991
82	TwdlH	8.94E-09	2.687
83	CG14569	8.99E-09	5.362
84	Cyp6w1	1.34E-08	2.553
85	CR43696	1.68E-08	3.332
86	CR42530	1.70E-08	2.785
87	CR42490	1.94E-08	-7.927
88	CR40190	1.94E-08	-1.117
89	Syp	2.29E-08	-2.491
90	Cib2	2.61E-08	-1.494
91	CG4259	2.75E-08	2.579
92	CG7408	3.48E-08	-1.753
93	CG14907	3.63E-08	1.524
94	yellow-d2	3.77E-08	-2.721
95	FucTC	3.95E-08	2.204
96	Ugt302E1	4.39E-08	3.723
97	Cyp28d1	4.59E-08	2.421
98	Cyp12e1	4.88E-08	-2.440
99	lncRNA:CR44603	4.93E-08	-2.915
100	CG7054	5.55E-08	1.663

Supplementary Table 1.5: Top 100 most DE genes with the lowest padj values (T14)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.82E-60	-6.716
2	w	9.75E-44	-4.637
3	Hsc70-2	2.67E-42	-3.921
4	TwdlU	6.94E-40	-5.793
5	CG42822	4.80E-34	5.095
6	lncRNA:CR43459	1.03E-32	-2.766
7	NLaz	4.58E-31	-4.365
8	CG11236	1.43E-30	-3.283
9	CG18547	2.50E-27	2.148
10	yellow-d2	6.57E-25	-4.974
11	ppk	8.04E-25	4.013
12	CG2233	8.04E-25	4.474
13	lncRNA:CR44536	7.13E-24	9.017
14	CG10445	3.36E-22	4.368
15	asRNA:CR44029	5.31E-22	-3.715
16	CR42490	9.69E-22	-6.943
17	CG31706	1.50E-21	4.424
18	CG2064	5.51E-21	2.877
19	CG3165	5.59E-21	2.538
20	CG8541	6.63E-20	3.854
21	CG14014	2.68E-19	3.088
22	CG10407	3.49E-19	-2.451
23	DNApol-zeta	1.10E-18	2.379
24	CG1494	1.43E-18	-2.965
25	Cyp12e1	1.95E-18	-3.423
26	CG18278	1.42E-17	3.779
27	CG11319	1.94E-17	-1.867
28	CG30059	4.36E-17	3.814
29	Cyp6a18	9.78E-17	-2.540
30	lncRNA:CR44466	1.91E-16	-3.358
31	Syp	3.22E-16	-3.536
32	Ect3	6.74E-16	-2.259
33	CG14406	4.28E-15	-3.237
34	CG13694	8.41E-15	2.859
35	CG42365	1.67E-14	-2.886
36	mthl3	1.73E-14	2.450
37	CG5687	1.73E-14	3.116
38	asRNA:CR42547	6.19E-14	-3.249
39	CG4730	6.19E-14	2.894
40	asRNA:CR45835	6.65E-14	-5.278
41	Odc1	7.64E-14	3.151
42	Nop17l	1.12E-13	-1.123
43	CG3397	1.15E-13	-4.026
44	CG31157	1.15E-13	2.845
45	CG9360	2.73E-13	-3.714
46	Cpr76Bc	3.13E-13	-2.908
47	CR43105	3.42E-13	5.113
48	Grik	3.62E-13	-3.149
49	SecCl	5.44E-13	-1.752
50	CG14456	6.46E-13	-2.082
51	Hmgcl	6.69E-13	-1.928
52	Drsl5	6.69E-13	4.188
53	asRNA:CR45171	6.69E-13	10.097
54	Sid	7.06E-13	-2.695
55	CG13907	8.21E-13	-1.298
56	lncRNA:CR46006	8.70E-13	-1.999
57	CG5171	1.13E-12	-2.900
58	CG31676	1.13E-12	-1.817
59	CG4842	1.87E-12	2.560
60	ms(3)76Ba	2.14E-12	9.278
61	CG6300	2.34E-12	4.618
62	Rev7	2.54E-12	2.282
63	agt	2.54E-12	2.728
64	Ugt302E1	3.14E-12	3.932
65	asRNA:CR45822	3.28E-12	-8.954
66	asRNA:CR45140	4.90E-12	-3.087
67	sni	5.07E-12	1.246
68	Bin1	5.29E-12	1.846
69	CG3842	5.51E-12	-1.829
70	CG13813	5.68E-12	-4.232
71	SCAP	7.47E-12	1.784
72	tut	7.61E-12	3.504
73	Ugt35B1	1.15E-11	-1.962
74	CG33337	1.16E-11	-3.781
75	narya	1.49E-11	6.086
76	Act87E	1.83E-11	-1.336
77	CG7054	2.53E-11	1.961
78	CG10799	3.02E-11	-2.064
79	Ir76b	3.02E-11	3.157
80	CG32369	3.62E-11	-1.366
81	alpha-Est2	4.13E-11	-1.789
82	asRNA:CR44030	5.35E-11	2.940
83	CG14204	1.04E-10	-4.008
84	CG5961	1.32E-10	-1.379
85	CR42530	1.46E-10	3.179
86	CG6293	1.65E-10	2.898
87	CG14356	1.95E-10	-2.539
88	VhaM9.7-a	3.09E-10	-1.475
89	Cpr11A	3.09E-10	2.485
90	CG10910	3.27E-10	4.034
91	CG33301	3.35E-10	5.327
92	CG12766	3.69E-10	2.856
93	Irbp	3.92E-10	2.248
94	FoxP	6.29E-10	-1.991
95	lncRNA:CR44754	8.09E-10	-2.209
96	CG14823	8.63E-10	-3.247
97	fusl	8.93E-10	-1.806
98	Syx4	1.31E-09	-1.472
99	CG11318	1.44E-09	2.390
100	rad50	1.61E-09	1.972

Supplementary Table 1.6: Top 100 most DE genes with the lowest padj values (T15)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	3.14E-62	-7.362
2	CG2233	9.47E-39	4.925
3	w	4.31E-37	-4.152
4	ppk	1.16E-34	5.430
5	CG42822	1.47E-33	4.567
6	NLaz	8.91E-28	-4.137
7	lncRNA:CR43459	8.91E-28	-2.559
8	CG18547	2.30E-27	2.129
9	Hsc70-2	4.40E-27	-2.961
10	CG2064	1.67E-26	3.231
11	CG11236	1.07E-23	-2.665
12	yellow-d2	3.50E-23	-4.323
13	CG11659	3.50E-23	5.851
14	CG6300	9.64E-23	6.128
15	CG31706	2.03E-22	3.610
16	CG6163	6.24E-22	-4.434
17	CG1494	7.70E-22	-3.147
18	Cyp12e1	2.11E-21	-3.558
19	asRNA:CR44029	5.52E-21	-3.577
20	CG12766	1.20E-20	3.981
21	CG10560	2.64E-20	4.066
22	CG12643	8.39E-20	2.912
23	CG8541	8.89E-20	3.389
24	CG17549	2.82E-19	3.691
25	Drsl5	6.15E-19	7.234
26	Bin1	1.02E-18	2.310
27	CG32241	1.52E-18	6.256
28	CG14406	2.30E-18	-2.918
29	Ugt302E1	1.14E-17	6.677
30	CG7299	2.13E-17	4.084
31	Jon74E	2.23E-17	4.819
32	CG11413	2.82E-17	3.856
33	SCAP	3.12E-17	2.151
34	asRNA:CR45835	7.48E-17	-4.280
35	CG11852	2.17E-16	5.811
36	TwdlU	2.60E-16	-4.116
37	yip2	4.21E-16	-1.959
38	CG6910	6.89E-16	5.494
39	CG10799	1.04E-15	3.843
40	CG40198	1.15E-15	5.176
41	CG15597	2.11E-15	3.966
42	Ugt37A3	2.78E-15	4.220
43	asRNA:CR45822	3.33E-15	-8.007
44	CG31157	8.36E-15	-4.240
45	CG11318	9.49E-15	2.774
46	CG3165	1.06E-14	2.134
47	CG10407	2.48E-14	-2.103
48	CG17191	2.84E-14	5.388
49	Grik	2.87E-14	-3.254
50	asRNA:CR44030	3.28E-14	3.610
51	Odc1	4.86E-14	2.976
52	CR43105	4.86E-14	4.164
53	CR43697	5.87E-14	2.739
54	sni	5.93E-14	1.337
55	Hmgcl	7.72E-14	-2.001
56	Cyp6a18	9.77E-14	-2.210
57	Nop17l	1.04E-13	-1.119
58	CG10445	1.13E-13	3.486
59	CG31103	1.25E-13	3.831
60	lncRNA:CR44536	1.30E-13	10.088
61	CG10513	2.22E-13	3.062
62	CG33301	2.95E-13	4.895
63	CG7054	4.13E-13	2.099
64	CG31104	8.14E-13	4.452
65	CG10638	1.09E-12	1.804
66	CG31288	1.09E-12	2.409
67	Ect3	1.45E-12	-2.004
68	Cyp6a13	1.60E-12	1.486
69	Cpr67B	2.66E-12	4.077
70	CG17738	4.23E-12	4.400
71	CG31098	4.33E-12	1.612
72	CG31778	5.22E-12	2.321
73	yellow-d	7.29E-12	2.886
74	CG3397	8.07E-12	2.471
75	Mdr50	9.33E-12	2.801
76	CG14694	1.05E-11	-3.269
77	CR45470	1.09E-11	4.766
78	asRNA:CR44960	1.11E-11	5.073
79	hpRNA:CR33940	1.21E-11	-2.956
80	Cyp12a4	1.21E-11	4.138
81	CG14204	1.26E-11	-4.606
82	CR40190	1.54E-11	-1.290
83	Gbp3	1.59E-11	2.571
84	CG31676	1.64E-11	-1.716
85	Cpr11B	1.73E-11	4.414
86	CG12825	1.89E-11	2.072
87	lncRNA:CR44466	2.22E-11	-2.732
88	Rev7	2.68E-11	2.144
89	Syp	2.99E-11	-3.072
90	Cyp6w1	3.14E-11	2.712
91	CG11319	4.12E-11	-1.467
92	asRNA:CR45140	4.36E-11	-2.855
93	lncRNA:CR45039	4.46E-11	4.151
94	CG9682	4.68E-11	5.290
95	CG5767	5.55E-11	6.039
96	CG9877	8.73E-11	5.528
97	Cht5	9.46E-11	2.806
98	DNApol-zeta	1.37E-10	1.779
99	CG34195	1.59E-10	-1.735
100	mthl3	1.71E-10	2.501

Supplementary Table 1.7: Top 100 most DE genes with the lowest padj values (T16)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	6.77E-73	-6.334
2	CG14406	1.21E-51	-5.214
3	NLaz	9.70E-38	-4.814
4	w	1.26E-34	-4.111
5	CG1494	3.89E-34	-3.971
6	CG11236	1.66E-30	-3.128
7	yellow-d2	4.27E-30	-5.293
8	ppk	1.13E-28	5.242
9	Hsc70-2	1.30E-28	-3.164
10	Ect3	6.90E-27	-2.954
11	CG11319	6.02E-26	-2.249
12	Grik	3.75E-25	-4.495
13	CG14661	4.81E-25	-3.295
14	Ugt302E1	2.06E-24	4.851
15	CG42807	1.87E-23	-3.159
16	lncRNA:CR43459	2.02E-23	-2.361
17	CG9864	8.53E-23	-3.256
18	CG30059	3.96E-22	4.812
19	CG3690	8.63E-22	-3.520
20	asRNA:CR44029	1.14E-21	-3.694
21	CG6296	2.68E-21	-3.182
22	CG3165	6.88E-21	2.550
23	CG18278	7.69E-21	4.226
24	CG12643	1.45E-20	2.974
25	Bin1	1.80E-20	2.424
26	CG13159	4.47E-20	-5.625
27	CG6300	4.47E-20	5.660
28	Ugt35B1	4.50E-20	-2.457
29	CG3819	7.50E-20	-5.152
30	Cyp12e1	8.70E-20	-3.418
31	CG11659	2.06E-19	5.154
32	Cyp6a18	9.15E-19	-2.597
33	lncRNA:CR44754	9.15E-19	-2.808
34	CG13813	1.05E-18	-5.335
35	asRNA:CR45835	1.22E-18	-5.646
36	MFS14	2.96E-18	-1.893
37	CG10445	5.96E-18	3.832
38	CG10407	6.82E-18	-2.323
39	CG14456	1.53E-17	-2.137
40	CG14694	1.53E-17	-4.796
41	tut	3.26E-17	4.355
42	CG42822	7.27E-17	3.285
43	CR42490	2.30E-16	-6.495
44	dpr21	3.39E-16	-2.111
45	Irbp	4.26E-16	2.855
46	Btnd	7.09E-16	-3.184
47	CG31676	7.59E-16	-2.036
48	CG2064	7.59E-16	2.497
49	Act87E	1.93E-15	-1.549
50	lncRNA:CR46006	1.98E-15	-2.190
51	lncRNA:CR44536	3.41E-15	10.068
52	DNApol-zeta	4.03E-15	2.128
53	CG10232	4.87E-15	-2.513
54	CG1304	5.64E-15	-5.135
55	CG11413	6.23E-15	-3.229
56	CG11000	1.07E-14	-1.682
57	Tmem18	1.13E-14	2.318
58	SecCl	1.34E-14	-1.804
59	sni	1.34E-14	1.349
60	Ir76b	1.40E-14	-2.231
61	CG14823	2.37E-14	-3.826
62	asRNA:CR44030	2.37E-14	3.475
63	Syt12	2.58E-14	-1.787
64	Drsl5	2.63E-14	6.305
65	yip2	4.40E-14	-1.824
66	Lk	4.71E-14	-3.938
67	Ssrp	4.77E-14	1.998
68	Fen1	5.97E-14	2.349
69	MED28	7.87E-14	1.747
70	CG32104	9.55E-14	1.708
71	CG32243	1.03E-13	2.615
72	CG34276	1.57E-13	-2.777
73	Nup43	1.57E-13	1.995
74	CG5171	1.58E-13	-2.997
75	Ugt35E1	1.79E-13	-2.488
76	NaPi-T	2.08E-13	-2.765
77	Mes4	2.60E-13	2.293
78	CG12863	3.61E-13	2.076
79	Nmdar1	3.61E-13	-1.688
80	CG5961	3.63E-13	-1.511
81	CG7300	3.66E-13	-3.372
82	CG1239	4.95E-13	3.334
83	CG32815	5.66E-13	-1.835
84	Obp56b	6.13E-13	-3.424
85	asRNA:CR42547	8.26E-13	-2.837
86	CG17977	1.13E-12	1.984
87	CG7011	1.48E-12	1.273
88	Hira	1.69E-12	1.995
89	btl	1.94E-12	-2.348
90	CG3270	2.27E-12	-2.874
91	JMJD7	2.34E-12	2.335
92	Obp99a	3.06E-12	-2.854
93	RunxA	3.24E-12	-2.467
94	Mat1	4.06E-12	2.213
95	alpha-Est2	4.16E-12	-1.837
96	CG8664	4.55E-12	-3.009
97	Rev7	4.66E-12	2.263
98	Oseg4	4.78E-12	-2.044
99	kar	4.81E-12	-1.924
100	CG2233	4.83E-12	2.534

Supplementary Table 1.8: Top 100 most DE genes with the lowest padj values (T17)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	8.84E-61	-5.974
2	CG14406	1.66E-51	-5.532
3	CG1494	1.97E-36	-4.178
4	w	9.15E-34	-4.408
5	ppk	5.47E-32	5.556
6	asRNA:CR44179	7.99E-32	-3.352
7	lncRNA:CR43459	5.23E-31	-2.738
8	Ugt302E1	4.22E-28	5.547
9	NLaz	6.63E-25	-3.936
10	yellow-d2	6.63E-25	-5.902
11	CG11319	6.66E-25	-2.208
12	lncRNA:CR44536	6.66E-25	9.201
13	CG11236	4.65E-24	-2.821
14	Hsc70-2	2.29E-23	-2.878
15	CG3165	1.04E-22	2.646
16	lncRNA:CR44754	1.10E-21	-3.276
17	CG14961	1.99E-21	-4.058
18	Cyp6a18	1.18E-20	-2.834
19	CG12643	2.92E-20	2.962
20	asRNA:CR44029	5.47E-20	-3.570
21	CG4730	6.13E-20	3.504
22	Grik	1.58E-19	-4.049
23	Ect3	1.78E-19	-2.535
24	CG42807	2.32E-19	-2.593
25	CG14661	4.04E-19	-2.744
26	CG11000	2.33E-18	-1.891
27	CG30091	5.08E-18	-4.205
28	CG6296	1.71E-17	-2.863
29	CG10407	1.98E-17	-2.302
30	dpr21	2.55E-17	-2.315
31	Syt12	2.58E-17	-2.003
32	CG10445	7.85E-17	3.799
33	CG9864	1.00E-16	-2.781
34	CG6398	3.92E-16	1.248
35	asRNA:CR45822	4.46E-16	-6.959
36	asRNA:CR44030	5.40E-16	4.041
37	CG3690	6.19E-16	-3.069
38	Bin1	8.20E-16	2.129
39	CG13813	1.27E-15	-4.985
40	CG5316	1.31E-15	1.366
41	CG1304	1.44E-15	-5.253
42	Lk	1.49E-15	-4.166
43	CG33128	1.68E-15	-5.267
44	CG5961	1.94E-15	-1.653
45	CG31676	2.87E-15	-2.016
46	CG32815	3.19E-15	-2.013
47	CG30059	4.29E-15	3.389
48	asRNA:CR45835	6.31E-15	-5.431
49	Cyp12e1	9.26E-15	-2.960
50	Drsl5	1.04E-14	6.384
51	CG18278	1.54E-14	3.271
52	Ugt37A3	1.54E-14	3.575
53	CG2233	1.77E-14	2.807
54	Nop17l	1.78E-14	-1.151
55	Nmdar1	1.78E-14	-1.779
56	Obp56b	2.02E-14	-3.881
57	CG14694	2.53E-14	-3.609
58	Ir76b	3.02E-14	-2.214
59	pigs	3.49E-14	-2.042
60	CG3819	4.85E-14	-4.335
61	Ugt35E1	7.08E-14	-2.536
62	CG6300	7.52E-14	4.575
63	CG11659	8.02E-14	4.223
64	lncRNA:CR31781	1.46E-13	-2.027
65	CG42822	1.53E-13	2.987
66	CR43105	2.18E-13	3.501
67	tut	2.43E-13	3.554
68	side	2.43E-13	-1.799
69	Btnd	2.59E-13	-2.921
70	lncRNA:CR46006	4.32E-13	-2.036
71	CG3552	5.60E-13	1.791
72	ZnT77C	6.38E-13	-1.687
73	ms(3)76Ba	7.07E-13	9.471
74	bma	1.04E-12	-1.718
75	CG3270	1.09E-12	-2.911
76	NaPi-T	1.97E-12	-2.664
77	Orcokinin	2.11E-12	-2.417
78	Cyp4ac3	2.47E-12	-3.043
79	narya	2.49E-12	6.259
80	CG10232	3.81E-12	-2.266
81	CR43697	5.72E-12	2.890
82	CG2064	8.03E-12	2.146
83	Oseg4	8.03E-12	-2.034
84	MFS14	8.23E-12	-1.525
85	CG6163	8.40E-12	-3.046
86	SecCl	8.78E-12	-1.619
87	Hug	9.10E-12	-3.111
88	CG14234	1.17E-11	-2.232
89	GlyT	1.22E-11	-2.175
90	RunxA	1.22E-11	-2.529
91	lncRNA:CR45039	1.27E-11	4.894
92	CG14456	1.86E-11	-1.711
93	CG13743	2.00E-11	-1.533
94	asRNA:CR42547	2.05E-11	-2.876
95	CG33337	2.11E-11	-3.007
96	CG42540	2.31E-11	-1.738
97	Ugt35B1	2.60E-11	-1.905
98	Capa	2.77E-11	-3.592
99	CR42530	2.78E-11	3.251
100	Pxd	2.88E-11	-1.608

Supplementary Table 1.9: Top 100 most DE genes with the lowest padj values (T18)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	2.66E-50	-6.129
2	CG14406	2.77E-49	-6.271
3	CG42807	3.27E-42	-3.253
4	CG3690	3.97E-38	-3.665
5	lncRNA:CR43459	7.70E-38	-3.024
6	CG6296	6.77E-37	-4.149
7	w	1.44E-33	-4.518
8	Ugt302E1	5.07E-33	5.623
9	CG1494	2.43E-32	-3.896
10	CG11236	3.00E-32	-3.340
11	CG3819	3.16E-30	-5.478
12	NaPi-T	2.84E-29	-3.939
13	CG11319	2.38E-28	-2.340
14	ppk	3.10E-28	4.726
15	asRNA:CR45822	3.96E-28	-6.521
16	lncRNA:CR44536	8.80E-27	8.700
17	Btnd	3.28E-26	-3.922
18	Ugt35E1	5.30E-26	-3.347
19	Ugt35B1	1.45E-25	-3.129
20	lncRNA:CR44754	3.99E-25	-3.800
21	dpr21	4.57E-24	-2.574
22	AcCoAS	1.91E-23	-2.102
23	CG3165	7.64E-22	2.613
24	Bin1	8.05E-22	2.519
25	Hsc70-2	3.84E-21	-2.807
26	Cyp6a18	4.10E-21	-2.885
27	CG11000	4.37E-21	-2.007
28	CG5961	6.43E-21	-1.908
29	Grik	6.43E-21	-4.326
30	asRNA:CR45835	1.09E-20	-4.422
31	Syt12	1.80E-20	-2.172
32	Obp99d	2.20E-20	-3.567
33	CG1304	4.86E-20	-5.925
34	Ssrp	4.89E-20	2.392
35	asRNA:CR44029	2.14E-19	-3.568
36	CG18278	4.96E-19	4.082
37	CG14661	1.33E-18	-2.620
38	CG30059	1.50E-18	4.064
39	sni	2.41E-18	1.521
40	ZnT77C	3.38E-18	-2.012
41	Mal-A8	3.66E-18	-3.053
42	Pxd	4.46E-18	-2.104
43	Irbp	7.88E-18	3.020
44	MED28	7.88E-18	2.025
45	lncRNA:CR31781	9.44E-18	-2.315
46	CG10445	1.31E-17	3.956
47	CG6733	1.56E-17	-3.337
48	DNApol-zeta	1.64E-17	2.296
49	CG33337	1.64E-17	-3.663
50	beta4GalT7	2.54E-17	2.923

Rank	Gene Symbol	padj value	log2 fc
51	CG30026	2.85E-17	-2.412
52	MFS14	3.01E-17	-1.833
53	CG3270	3.19E-17	-3.350
54	RpA-70	3.94E-17	2.848
55	beta4GalNAcTA	5.22E-17	-1.670
56	Mes4	5.23E-17	2.677
57	CG12643	8.07E-17	2.664
58	SmD3	8.99E-17	2.503
59	Usp5	1.24E-16	2.505
60	lncRNA:CR46006	1.34E-16	-2.303
61	CG9826	1.56E-16	-3.086
62	CG18327	1.70E-16	-2.971
63	CG31676	1.73E-16	-2.083
64	CG9344	1.94E-16	2.587
65	CG31249	1.94E-16	1.946
66	CG5246	2.10E-16	-3.803
67	CG32104	3.03E-16	1.903
68	CG8745	4.01E-16	-3.120
69	Orc4	1.42E-15	2.747
70	Gpb5	1.51E-15	-2.961
71	MED16	1.82E-15	2.130
72	Cpsf73	2.10E-15	2.197
73	Psf3	3.47E-15	2.682
74	CR40190	4.34E-15	-1.484
75	alpha-Est2	4.34E-15	-2.050
76	SecCl	4.35E-15	-1.796
77	Orcokinin	4.54E-15	-2.524
78	DIP-lambda	4.84E-15	-1.938
79	CG9864	5.44E-15	-2.600
80	yellow-d2	5.71E-15	-5.588
81	Ugt37A3	5.71E-15	3.563
82	Obp56b	1.07E-14	-3.466
83	CG32581	1.12E-14	-1.324
84	CG10407	1.20E-14	-2.066
85	SCOT	1.21E-14	-1.380
86	Cyp12e1	1.23E-14	-2.980
87	nAChRalpha7	1.25E-14	-1.932
88	JMJD7	1.60E-14	2.548
89	Ect3	1.92E-14	-2.165
90	Ir76b	2.36E-14	-2.232
91	Gal	2.66E-14	-2.195
92	CG10361	2.83E-14	-2.367
93	tut	3.07E-14	3.793
94	beta4GalNAcTB	3.35E-14	1.762
95	CG31191	3.35E-14	-1.597
96	CG32302	4.47E-14	-2.939
97	NijC	4.63E-14	-1.963
98	daw	5.06E-14	-2.050
99	Cdc23	5.60E-14	2.739
100	Act87E	5.60E-14	-1.463

Supplementary Table 1.10: Top 100 most DE genes with the lowest padj values (T19)

Rank	Gene Symbol	padj value	log2 fc
1	CG3819	1.21E-47	-6.621
2	CG14406	1.03E-41	-6.876
3	mthl8	2.89E-41	-6.721
4	CG3690	1.15E-36	-3.271
5	CG1494	6.61E-35	-4.463
6	lncRNA:CR43459	3.02E-30	-2.721
7	CG11236	7.95E-30	-3.148
8	w	6.57E-28	-3.843
9	CG42807	1.82E-27	-2.519
10	ppk	2.38E-26	5.295
11	CG5961	2.53E-26	-2.143
12	asRNA:CR45822	2.88E-26	-6.754
13	Ugt302E1	3.67E-26	7.029
14	NaPi-T	4.35E-26	-3.654
15	CG11319	6.23E-25	-2.196
16	Hsc70-2	6.89E-25	-3.035
17	CG12643	1.77E-23	3.204
18	Bin1	3.12E-23	2.614
19	lncRNA:CR44536	4.57E-23	8.965
20	lncRNA:CR44754	2.47E-22	-3.478
21	CG6733	3.38E-22	-3.782
22	CG33337	1.12E-21	-4.040
23	CG6296	8.29E-21	-2.976
24	CG10407	1.79E-20	-2.478
25	CG2064	1.10E-19	2.803
26	Btnd	3.27E-19	-3.311
27	Ugt35E1	4.92E-19	-2.773
28	nop5	5.22E-19	-0.895
29	CG3270	5.83E-19	-3.510
30	CG1304	7.76E-19	-5.757
31	AcCoAS	1.78E-18	-1.869
32	Grik	1.02E-17	-3.859
33	CG10638	2.13E-17	2.154
34	asRNA:CR44029	2.56E-17	-3.331
35	Ugt35B1	3.60E-17	-2.545
36	CG18327	8.69E-17	-2.927
37	asRNA:CR45835	2.10E-16	-5.034
38	MFS14	2.24E-16	-1.795
39	bnk	2.45E-16	2.859
40	CG10445	2.46E-16	4.214
41	Obp99d	2.64E-16	-3.065
42	CG3165	5.02E-16	2.244
43	Nop17l	7.53E-16	-1.205
44	Art8	8.24E-16	-1.530
45	CG3397	1.07E-15	3.363
46	Mal-A8	2.30E-15	-2.849
47	CG3902	2.42E-15	-1.940
48	Ku80	3.20E-15	2.969
49	GstD5	3.71E-15	5.036
50	CG9826	4.36E-15	-2.909

Rank	Gene Symbol	padj value	log2 fc
51	CG5402	4.79E-15	-3.256
52	CG11000	4.79E-15	-1.693
53	Ect3	5.29E-15	-2.239
54	Syt12	1.17E-14	-1.829
55	CG17219	1.65E-14	-1.752
56	CG5246	1.65E-14	-3.238
57	Gpb5	2.26E-14	-3.008
58	CG14014	3.67E-14	2.166
59	CG8745	5.86E-14	-2.926
60	asRNA:CR44030	9.44E-14	4.200
61	Gal	1.07E-13	-2.157
62	CG15784	1.08E-13	3.543
63	Mccc2	1.42E-13	-1.816
64	dpr21	1.69E-13	-1.882
65	Lk	1.72E-13	-3.495
66	Gba1a	1.72E-13	-2.540
67	CG9799	2.84E-13	-1.575
68	lncRNA:CR44466	3.49E-13	-3.594
69	Orcokinin	3.85E-13	-2.338
70	CG33124	4.80E-13	-2.607
71	RpA-70	5.03E-13	2.496
72	lncRNA:CR31781	5.82E-13	-1.982
73	CG4730	8.23E-13	2.848
74	ACC	9.72E-13	-1.795
75	DAT	1.01E-12	-2.120
76	beta4GalNAcTA	1.20E-12	-1.445
77	NHP2	1.20E-12	-1.274
78	FASN1	1.24E-12	-2.310
79	Cyp6a18	1.24E-12	-2.204
80	Ugt37A3	1.49E-12	3.232
81	Uhg2	1.60E-12	-1.823
82	Alg9	1.78E-12	2.649
83	CG18278	2.06E-12	3.225
84	CG14823	2.30E-12	-3.859
85	CG1671	2.34E-12	-1.519
86	CG9344	2.37E-12	2.394
87	Gclc	2.50E-12	-1.357
88	CG11659	3.20E-12	3.923
89	l(1)G0020	3.65E-12	-1.543
90	CG7966	3.91E-12	-1.864
91	CG5599	3.91E-12	-1.422
92	CG11158	4.56E-12	-1.928
93	CG6638	4.83E-12	-1.471
94	ZnT77C	4.83E-12	-1.633
95	tut	5.12E-12	3.709
96	NijC	7.64E-12	-1.924
97	CG8064	8.80E-12	-1.448
98	CG6770	9.27E-12	1.448
99	CG3552	9.71E-12	1.782
100	CG6300	9.71E-12	4.169

Supplementary Table 1.11: Top 100 most DE genes with the lowest padj values (T20)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	5.72E-41	-6.948
2	Ugt302E1	5.72E-41	7.413
3	CG3819	5.16E-40	-5.936
4	lncRNA:CR44536	1.23E-38	8.758
5	lncRNA:CR43459	6.03E-33	-2.857
6	ppk	8.94E-33	5.586
7	Hsc70-2	1.04E-31	-3.384
8	CG12643	1.88E-31	3.675
9	CG2064	1.63E-27	3.308
10	CG3690	2.90E-27	-2.703
11	Obp99d	4.94E-25	-3.749
12	w	2.52E-24	-3.523
13	CG14406	1.03E-20	-8.079
14	CG34166	1.57E-20	3.678
15	nop5	2.22E-20	-0.931
16	CG34195	5.16E-20	-2.447
17	CG33337	1.94E-19	-3.949
18	Cyp6a17	5.09E-19	2.720
19	asRNA:CR45835	5.68E-18	-4.098
20	Odc2	2.21E-17	3.837
21	Grik	4.44E-17	-3.864
22	CG3270	2.04E-16	-3.294
23	Ugt35B1	2.35E-16	-3.095
24	Cyp28d2	4.11E-16	2.694
25	CG5961	6.63E-16	-1.670
26	Ugt37A1	7.47E-16	4.062
27	CG11236	7.67E-16	-2.208
28	CG18327	1.16E-15	-2.836
29	CG10474	1.25E-15	-5.445
30	GstD5	1.48E-15	5.108
31	CG1494	1.85E-15	-2.710
32	mrt	2.29E-15	1.598
33	CG1304	3.74E-15	-5.187
34	asRNA:CR44030	2.10E-14	4.628
35	CG5402	2.43E-14	-3.335
36	NaPi-T	2.51E-14	-2.698
37	CG6770	2.81E-14	1.611
38	lncRNA:CR44754	2.85E-14	-3.193
39	CR43697	5.22E-14	3.489
40	aay	6.10E-14	-0.933
41	CG6296	8.04E-14	-2.414
42	CG31300	8.65E-14	2.842
43	CG5157	8.77E-14	2.072
44	CG17219	8.97E-14	-1.697
45	CG7458	1.06E-13	-1.264
46	CG6712	1.34E-13	-1.364
47	CG33124	1.47E-13	-2.699
48	lncRNA:CR44662	1.47E-13	3.419
49	CG3397	2.28E-13	2.673
50	Reg-2	2.63E-13	1.966
51	CG13694	3.59E-13	2.812
52	CG11319	3.88E-13	-1.596
53	CG10407	3.88E-13	-1.973
54	CG8654	4.11E-13	2.161
55	CG17977	6.07E-13	2.272
56	CG9515	6.11E-13	-1.375
57	U3-55K	6.17E-13	-1.517
58	Phae2	6.25E-13	2.575
59	Cat	6.25E-13	1.134
60	Mys45A	6.83E-13	-1.201
61	CG42807	6.83E-13	-1.686
62	Art8	8.03E-13	-1.379
63	Bin1	9.22E-13	1.925
64	CG6733	1.18E-12	-2.857
65	NHP2	1.41E-12	-1.274
66	CG13078	1.74E-12	4.605
67	firl	1.77E-12	-2.463
68	CG3552	2.99E-12	1.830
69	CG8665	3.50E-12	2.760
70	CG34105	3.72E-12	8.519
71	CG16775	6.25E-12	-3.840
72	asRNA:CR45822	6.27E-12	-8.667
73	CG4335	7.47E-12	3.166
74	CG10445	8.88E-12	3.212
75	CG42319	9.08E-12	2.159
76	CG11158	1.11E-11	-1.906
77	PH4alphaMP	1.30E-11	-2.505
78	CG6654	2.05E-11	1.479
79	CG6763	2.22E-11	-2.756
80	Fit2	2.52E-11	-1.347
81	CG13003	2.65E-11	-2.104
82	l(2)09851	3.43E-11	-1.056
83	Lk	3.52E-11	-3.210
84	Ugt37A3	3.91E-11	3.015
85	CG8064	4.09E-11	-1.412
86	Cyp4e3	4.10E-11	4.746
87	asRNA:CR44029	6.19E-11	-2.607
88	lncRNA:CR46006	6.64E-11	-1.878
89	DNApol-zeta	6.68E-11	1.799
90	CG3165	7.00E-11	1.844
91	CG30414	7.08E-11	2.257
92	CG10175	8.09E-11	-1.962
93	CG31683	9.29E-11	1.207
94	Drsl2	1.00E-10	6.622
95	CG15784	1.02E-10	3.123
96	lncRNA:CR45039	1.29E-10	5.578
97	lncRNA:CR43496	1.62E-10	2.260
98	Nop56	1.62E-10	-1.062
99	Cpr51A	1.70E-10	-2.729
100	Prat2	1.91E-10	2.114

Supplementary Table 1.12: Top 100 most DE genes with the lowest padj values (T21)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.82E-52	-5.952
2	Hsc70-2	1.66E-43	-4.141
3	CG3819	7.77E-42	-6.100
4	CG3690	1.56E-34	-3.058
5	Ugt302E1	1.19E-33	6.930
6	asRNA:CR45822	2.74E-33	-6.185
7	ppk	1.58E-30	6.079
8	Obp99d	3.10E-29	-4.099
9	lncRNA:CR43459	1.26E-28	-2.648
10	CG12643	8.59E-27	3.403
11	CG2064	2.56E-26	3.234
12	lncRNA:CR44536	4.28E-26	9.532
13	CG14406	5.40E-26	-7.290
14	Cyp6a17	1.88E-24	3.090
15	w	2.11E-23	-3.522
16	Odc2	4.48E-22	4.641
17	Cyp28d2	1.47E-20	3.052
18	Grik	1.61E-20	-4.347
19	CG11319	2.47E-20	-1.989
20	asRNA:CR45835	5.63E-20	-4.483
21	CG42807	1.52E-19	-2.085
22	CG8665	1.93E-19	3.521
23	CG6398	8.38E-19	1.354
24	Nop17l	1.93E-18	-1.298
25	ms(3)76Ba	2.00E-18	7.810
26	CG1494	9.66E-18	-2.954
27	CG3165	4.60E-17	2.307
28	CG11236	3.13E-16	-2.247
29	aay	9.52E-16	-0.989
30	CG6296	1.22E-15	-2.567
31	Oseg4	2.61E-15	-2.348
32	CG5961	3.46E-15	-1.634
33	CG1304	8.80E-15	-5.127
34	Lk	2.17E-14	-3.824
35	CG13907	1.04E-13	-1.355
36	CG10407	1.20E-13	-2.002
37	CG5316	1.29E-13	1.294
38	asRNA:CR44030	1.35E-13	4.472
39	dpr21	2.42E-13	-1.929
40	CG3270	5.38E-13	-2.929
41	Bin1	8.79E-13	1.933
42	Gpb5	1.03E-12	-2.951
43	CG13962	1.04E-12	4.178
44	CG33337	1.16E-12	-3.140
45	CG10445	1.23E-12	3.339
46	CG4730	1.34E-12	2.825
47	GstD5	1.81E-12	4.571
48	CG34195	1.93E-12	-1.825
49	Ugt37A3	1.97E-12	3.224
50	Ade2	2.39E-12	-1.323
51	CG5599	3.17E-12	-1.439
52	lncRNA:CR45039	3.39E-12	4.659
53	CG42319	8.06E-12	2.182
54	CG31300	8.06E-12	2.608
55	lncRNA:CR44662	8.06E-12	3.155
56	Mal-A8	1.01E-11	-2.453
57	asRNA:CR45171	1.25E-11	4.244
58	CG31683	1.34E-11	1.257
59	NaPi-T	1.66E-11	-2.419
60	CG18858	1.79E-11	1.145
61	upSET	1.84E-11	1.504
62	Iyd	2.18E-11	2.843
63	asRNA:CR44029	2.30E-11	-2.653
64	CG2765	2.51E-11	-0.816
65	CG8654	2.82E-11	2.012
66	Mccc2	3.01E-11	-1.661
67	CG4335	3.23E-11	3.203
68	CG12896	4.98E-11	-3.256
69	CR43105	6.04E-11	2.957
70	CG10474	6.47E-11	-4.491
71	U3-55K	7.40E-11	-1.392
72	CG6638	8.52E-11	-1.405
73	CG34166	8.82E-11	2.657
74	CG11000	9.32E-11	-1.439
75	lncRNA:CR44754	1.07E-10	-2.466
76	CG16704	1.08E-10	3.248
77	CG30059	1.27E-10	3.030
78	CR43697	1.39E-10	2.799
79	CAH2	1.58E-10	-1.316
80	Cyp4e3	1.94E-10	3.764
81	Ugt36D1	1.95E-10	1.767
82	CG18327	2.32E-10	-2.265
83	Obp56b	3.18E-10	-2.820
84	CG34215	3.82E-10	2.217
85	lectin-24A	3.88E-10	3.816
86	CG10932	3.96E-10	-1.807
87	CG17977	4.99E-10	1.942
88	l(2)09851	7.80E-10	-0.992
89	CG8064	8.34E-10	-1.330
90	Cpr64Ac	1.07E-09	-6.785
91	CG31157	1.07E-09	-3.214
92	Cyp12d1-d	1.45E-09	2.555
93	CG6665	1.72E-09	1.214
94	mrt	1.76E-09	1.252
95	CG10638	1.78E-09	1.578
96	CG6770	1.84E-09	-3.912
97	CG13159	1.84E-09	1.307
98	CG15432	2.18E-09	1.123
99	Dph3	2.31E-09	-1.121
100	CG5402	2.35E-09	-2.595

Supplementary Table 1.13: Top 100 most DE genes with the lowest padj values (T22)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	5.83E-55	-7.742
2	Ugt302E1	1.37E-54	6.009
3	ppk	9.95E-44	6.363
4	CG3819	3.37E-37	-5.676
5	lncRNA:CR44536	2.59E-36	8.895
6	Grik	4.63E-34	-5.493
7	CG14406	4.17E-31	-8.008
8	CG9360	8.11E-31	-5.666
9	Cyp28d2	6.82E-30	3.647
10	CG10407	1.08E-29	-2.884
11	firl	3.35E-29	-3.741
12	lncRNA:CR43459	2.46E-28	-2.626
13	CG33337	1.19E-27	-4.905
14	Hsc70-2	1.98E-27	-2.979
15	Obp99d	2.89E-27	-4.292
16	Odc2	3.59E-26	4.643
17	GstE1	6.56E-26	-3.279
18	Keap1	2.60E-24	-1.539
19	CG30091	7.34E-23	-4.497
20	CG12643	2.26E-22	3.033
21	CG14186	3.46E-22	-3.301
22	Cpr51A	5.12E-21	-3.889
23	w	5.12E-21	-3.251
24	asRNA:CR44030	1.05E-20	4.500
25	ZnT77C	1.68E-20	-2.124
26	asRNA:CR44179	6.32E-20	-2.722
27	CG6296	1.99E-18	-2.768
28	Ect3	2.86E-18	-2.468
29	CG17977	2.98E-18	2.419
30	CG32104	5.80E-18	1.888
31	CG9733	6.47E-18	-4.916
32	NLaz	7.56E-18	-3.322
33	asRNA:CR45835	1.12E-17	-5.173
34	CG2150	1.53E-17	-5.286
35	Oseg4	2.91E-17	-2.492
36	Obp99a	6.23E-17	-3.364
37	Phk-3	7.18E-17	-2.600
38	Cpr64Ac	7.18E-17	-9.004
39	CR43697	8.95E-17	3.778
40	Sb	1.08E-16	-4.441
41	CG42807	1.11E-16	-1.926
42	CG13159	1.20E-16	-5.140
43	CG5402	1.72E-16	-3.529
44	CG30026	3.61E-16	-2.349
45	TwdlX	4.24E-16	-2.194
46	CG43204	5.57E-16	-3.724
47	lncRNA:CR44754	5.57E-16	-3.252
48	CG11413	6.89E-16	-3.276
49	CG7406	6.89E-16	-6.665
50	CG31869	1.17E-15	-2.465
51	scaf	1.51E-15	-2.465
52	asRNA:CR45822	2.21E-15	-8.055
53	lncRNA:CR42646	3.25E-15	-4.595
54	CG34105	5.55E-15	10.276
55	CG14567	6.56E-15	-4.585
56	ChLD3	6.68E-15	-4.835
57	CG14961	8.08E-15	-3.266
58	CG7912	8.08E-15	2.676
59	Gadd45	8.91E-15	-3.039
60	CG7724	1.79E-14	-2.068
61	asRNA:CR45140	1.96E-14	-3.224
62	ThrRS	2.32E-14	-0.917
63	Dyrk2	3.09E-14	-1.271
64	Muc91C	3.09E-14	-6.742
65	CG13606	3.09E-14	-2.376
66	CG8654	3.53E-14	2.192
67	CG3552	4.34E-14	1.806
68	CG6654	4.38E-14	1.606
69	CG34437	4.82E-14	-3.220
70	Cpr35B	5.04E-14	-3.000
71	CG12863	5.04E-14	2.090
72	Cyp6a17	5.04E-14	2.276
73	Cpr65Ea	5.04E-14	-4.352
74	TwdlC	5.04E-14	-6.512
75	Lk	7.89E-14	-3.869
76	tut	9.93E-14	3.865
77	CG2064	1.00E-13	2.292
78	Cul2	1.07E-13	-2.179
79	CG12164	1.09E-13	-4.740
80	Mfap1	1.20E-13	-1.669
81	Rlip	1.35E-13	1.198
82	CG13063	1.74E-13	-4.915
83	CG11601	2.01E-13	1.421
84	CG15731	2.04E-13	-7.146
85	CG13403	2.35E-13	-2.691
86	dyl	2.99E-13	-6.455
87	CG2962	3.79E-13	-6.592
88	Trh	3.98E-13	-2.079
89	lncRNA:CR45039	3.98E-13	4.974
90	CG1494	3.98E-13	-2.425
91	CR45470	5.52E-13	4.728
92	CG3690	5.52E-13	-1.850
93	Nop17l	6.20E-13	-1.076
94	CG13699	6.28E-13	-5.671
95	ple	6.53E-13	-3.207
96	CG7300	9.45E-13	-3.176
97	Lectin-galC1	1.27E-12	3.439
98	Ugt317A1	1.53E-12	-2.552
99	Ccp84Ac	1.78E-12	-6.208
100	e	1.78E-12	-2.260

Supplementary Table 1.14: Top 100 most DE genes with the lowest padj values (T23)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.31E-71	-6.680
2	Ugt302E1	1.08E-41	4.851
3	ppk	2.91E-39	6.051
4	CG3819	4.20E-36	-5.567
5	lncRNA:CR43459	5.25E-35	-2.928
6	CG10407	1.44E-34	-3.113
7	Hsc70-2	2.69E-34	-3.593
8	Keap1	4.03E-32	-1.771
9	Grik	8.11E-31	-5.199
10	firl	2.83E-30	-3.748
11	ZnT77C	1.35E-29	-2.550
12	lncRNA:CR44536	7.07E-29	9.014
13	Odc2	2.19E-28	4.900
14	Cyp28d2	2.42E-28	3.559
15	CG14406	4.89E-27	-8.159
16	CG1494	2.56E-26	-3.638
17	Obp99d	6.07E-25	-3.966
18	CG12643	4.50E-23	3.096
19	GstE1	9.67E-23	-3.079
20	CG6296	1.50E-21	-2.984
21	Cpr51A	1.54E-21	-3.989
22	Phk-3	2.25E-21	-2.926
23	Cpr65Ea	5.55E-21	-5.436
24	asRNA:CR45140	2.17E-20	-3.898
25	CG7912	2.31E-20	3.167
26	mrt	3.26E-20	1.823
27	CG30091	1.99E-19	-4.036
28	CG33337	1.99E-19	-3.979
29	Ect3	1.29E-18	-2.490
30	CG31869	1.36E-18	-2.690
31	ThrRS	1.37E-18	-1.046
32	CG7724	1.40E-18	-2.352
33	CG9360	5.03E-18	-4.303
34	asRNA:CR45835	1.07E-17	-3.737
35	CG13159	1.30E-17	-5.306
36	Cyp6a23	1.30E-17	-2.424
37	Cul2	1.76E-17	-2.471
38	CG42807	5.11E-17	-1.941
39	CG30026	8.86E-17	-2.325
40	e	1.27E-16	-2.627
41	Spn100A	1.39E-16	-3.582
42	CG6654	1.47E-16	1.759
43	lncRNA:CR42646	1.76E-16	-4.892
44	asRNA:CR44179	2.77E-16	-2.376
45	CG43204	5.00E-16	-3.667
46	CG31810	6.25E-16	-2.786
47	CG12164	6.25E-16	-5.214
48	cysu	7.91E-16	-3.746
49	mRpS9	8.65E-16	0.901
50	ckd	1.34E-15	-2.190

Rank	Gene Symbol	padj value	log2 fc
51	Sb	1.77E-15	-4.290
52	w	1.93E-15	-2.735
53	CG34105	3.18E-15	10.417
54	CG15022	3.53E-15	-5.273
55	scaf	3.90E-15	-2.431
56	Spn43Aa	3.90E-15	-5.054
57	twz	4.47E-15	2.342
58	CG7227	4.62E-15	-2.900
59	ZnT35C	6.23E-15	2.010
60	lncRNA:CR45039	9.76E-15	4.980
61	CG43386	1.88E-14	-2.832
62	asRNA:CR44030	2.42E-14	3.659
63	asRNA:CR45822	2.97E-14	-7.730
64	Fnta	3.81E-14	0.998
65	CR43697	4.21E-14	3.464
66	Trh	5.01E-14	-2.143
67	CG14186	5.89E-14	-2.586
68	CG4783	5.89E-14	-6.475
69	CG31098	5.95E-14	1.725
70	CG17672	6.37E-14	-4.208
71	Cyp4p1	8.79E-14	-3.065
72	CG42319	1.14E-13	2.312
73	CG12769	1.42E-13	-2.550
74	Dyrk2	1.45E-13	-1.241
75	CG12491	1.45E-13	9.199
76	Vajk4	2.13E-13	-6.242
77	DIP-lambda	2.19E-13	-1.811
78	Ugt317A1	2.25E-13	-2.652
79	slbo	2.25E-13	-4.825
80	CG9733	2.25E-13	-4.027
81	CG2150	2.25E-13	-4.554
82	CG33346	2.50E-13	-4.160
83	CG33468	2.54E-13	-5.060
84	CG8654	2.54E-13	2.115
85	CG5402	2.54E-13	-3.017
86	CG34038	2.67E-13	-3.197
87	CG16935	2.68E-13	1.675
88	PPP1R15	2.96E-13	-0.963
89	CG13699	3.71E-13	-5.877
90	CG13659	3.71E-13	-2.983
91	asRNA:CR44029	4.51E-13	-2.822
92	CG17977	5.26E-13	2.174
93	CG8001	6.76E-13	1.362
94	CG4702	7.29E-13	-4.080
95	CG13813	7.38E-13	-5.568
96	Cad88C	7.39E-13	-2.806
97	CG10353	8.42E-13	1.688
98	CG33993	8.42E-13	-4.254
99	CG11147	8.52E-13	-2.565
100	Eip74EF	1.14E-12	-2.907

Supplementary Table 2.1: Top 100 downregulated genes with the lowest padj values (T10)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.48E-72	-5.415
2	w	5.49E-48	-4.729
3	mthl8	1.06E-47	-8.077
4	Grik	1.13E-42	-6.430
5	lncRNA:CR43459	2.70E-33	-2.815
6	ZnT77C	3.32E-30	-2.605
7	CG33337	4.52E-30	-8.391
8	CG11319	1.14E-27	-2.771
9	AstC-R2	1.00E-26	-3.354
10	CG14990	5.03E-25	-3.361
11	asRNA:CR45835	1.89E-24	-5.508
12	SecCl	1.89E-24	-2.422
13	CG9360	7.61E-24	-4.968
14	Cyp4p2	1.52E-23	-3.466
15	CR42490	1.88E-23	-7.155
16	CG31157	5.50E-22	-4.925
17	asRNA:CR45822	7.92E-22	-9.710
18	asRNA:CR44029	1.14E-21	-3.618
19	lncRNA:CR44466	2.59E-20	-3.699
20	dpr20	4.60E-20	-2.459
21	CG42365	1.30E-18	-3.244
22	CG11236	1.77E-17	-3.078
23	lncRNA:CR44754	1.89E-17	-3.561
24	Cyp12c1	5.98E-17	-3.330
25	mthl4	6.19E-17	-2.731
26	CG14301	6.19E-17	-2.345
27	CG32195	8.33E-17	-2.381
28	nAChRalpha7	1.86E-16	-2.410
29	mGluR	5.47E-16	-2.363
30	NLaz	1.36E-15	-3.099
31	CG14356	1.86E-15	-3.244
32	CG14204	2.30E-15	-4.167
33	CG31676	2.36E-15	-2.110
34	yellow-h	2.57E-15	-2.258
35	CG33468	3.50E-15	-3.941
36	nompC	4.03E-15	-1.919
37	St3	7.58E-15	-2.511
38	Scgalpha	1.17E-14	-2.292
39	CG31997	2.00E-14	-2.233
40	CG9568	2.51E-14	-1.949
41	Ect3	4.79E-14	-2.122
42	asRNA:CR45140	7.21E-14	-3.122
43	CG43338	1.04E-13	-1.946
44	Ugt35A1	1.23E-13	-1.607
45	CG42337	1.68E-13	-1.865
46	CG10407	1.89E-13	-1.964
47	lncRNA:CR46075	2.52E-13	-4.143
48	CR44383	4.78E-13	-5.397
49	CG34155	5.61E-13	-2.170
50	lncRNA:CR44743	6.75E-13	-3.970
51	Cyp9c1	8.83E-13	-2.575
52	Ugt302K1	1.30E-12	-6.782
53	CG34354	1.30E-12	-2.050
54	lncRNA:CR45127	2.04E-12	-6.788
55	CG4752	3.68E-12	-1.601
56	Cib2	4.04E-12	-2.107
57	CG34384	5.83E-12	-1.677
58	CG43155	8.23E-12	-2.520
59	Cngl	1.11E-11	-1.772
60	CG3842	1.23E-11	-1.783
61	CG33128	1.53E-11	-5.242
62	Ugt303B3	2.35E-11	-2.360
63	CG16799	2.54E-11	-1.943
64	yellow-d2	8.42E-11	-4.515
65	lncRNA:CR46006	8.43E-11	-1.787
66	dsb	8.95E-11	-3.181
67	lncRNA:CR44467	1.00E-10	-2.540
68	CG1494	1.08E-10	-2.240
69	CG6044	1.18E-10	-1.750
70	tgy	1.59E-10	-2.073
71	mmd	3.86E-10	-1.654
72	Had2	4.58E-10	-3.141
73	lncRNA:CR31781	5.08E-10	-1.757
74	unc79	5.73E-10	-1.561
75	Pxd	6.98E-10	-1.612
76	CG4409	7.16E-10	-1.917
77	Orcokinin	8.36E-10	-2.022
78	CG33521	1.18E-09	-1.897
79	Cyp6d4	1.30E-09	-2.341
80	FoxP	1.33E-09	-1.979
81	CG8768	1.35E-09	-1.718
82	CaMKI	1.85E-09	-1.473
83	antr	1.95E-09	-2.790
84	kek6	2.02E-09	-1.266
85	CR40190	2.04E-09	-1.133
86	lncRNA:CR45226	2.07E-09	-8.420
87	ttm2	2.18E-09	-3.545
88	Ugt302C1	2.48E-09	-1.264
89	CG15818	2.55E-09	-1.978
90	ATPsynCF6L	2.80E-09	-7.910
91	CG10051	2.96E-09	-1.863
92	Eip63F-1	3.04E-09	-1.740
93	GluRIIE	3.66E-09	-1.774
94	Argk	3.66E-09	-1.710
95	hpRNA:CR33940	3.71E-09	-2.540
96	CG15765	4.61E-09	-1.620
97	CG10232	5.26E-09	-2.058
98	Jheh3	6.04E-09	-1.660
99	CG11149	6.12E-09	-2.124
100	CG42566	8.77E-09	-1.420

Supplementary Table 2.2: Top 100 downregulated genes with the lowest padj values (T11)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.26E-70	-5.496
2	mthl8	6.82E-61	-6.800
3	w	9.73E-52	-4.932
4	Grik	1.13E-36	-5.499
5	CG33337	1.07E-30	-7.625
6	CR42490	9.96E-26	-7.018
7	ZnT77C	1.02E-24	-2.352
8	asRNA:CR45835	3.37E-24	-4.658
9	asRNA:CR45822	1.22E-22	-9.914
10	CG11236	1.25E-22	-4.285
11	lncRNA:CR44754	4.89E-22	-3.700
12	asRNA:CR44029	9.50E-21	-3.507
13	CG31157	5.07E-20	-4.540
14	lncRNA:CR43459	6.30E-20	-2.163
15	CG14204	7.41E-19	-5.443
16	CG9360	3.87E-17	-4.242
17	Cyp4p2	1.91E-16	-2.805
18	asRNA:CR45140	2.11E-16	-3.485
19	CG14301	3.59E-16	-2.354
20	Nop17l	1.18E-15	-1.194
21	lncRNA:CR46075	2.17E-15	-4.615
22	Ect3	2.56E-15	-2.213
23	CG31676	3.25E-15	-2.103
24	CG10407	5.71E-15	-2.072
25	Argk	1.08E-13	-2.096
26	mthl4	1.85E-13	-2.373
27	CR40190	1.86E-12	-1.303
28	lncRNA:CR44466	3.90E-12	-2.449
29	FoxP	4.58E-12	-2.238
30	CG15553	6.20E-12	-4.886
31	CG14990	9.42E-12	-2.159
32	Tina-1	9.90E-12	-1.200
33	Blos1	1.72E-11	-0.961
34	smp-30	2.17E-11	-2.513
35	Spn77Bc	2.19E-11	-3.199
36	NijC	2.19E-11	-1.705
37	lncRNA:CR45242	2.48E-11	-2.884
38	lncRNA:CR44743	4.47E-11	-3.456
39	CG14694	5.81E-11	-3.007
40	CG7607	1.13E-10	-2.259
41	lncRNA:CR44498	2.67E-10	-3.296
42	asRNA:CR42547	2.86E-10	-2.244
43	Jheh3	3.10E-10	-1.777
44	NLaz	3.39E-10	-2.487
45	CG33128	5.40E-10	-4.764
46	SecCl	8.58E-10	-1.492
47	CG12502	9.37E-10	-2.759
48	Ugt35A1	1.06E-09	-1.339
49	su(w[a])	1.12E-09	-0.828
50	asRNA:CR44053	1.25E-09	-2.177
51	Toll-9	1.38E-09	-2.232
52	ATPsynCF6L	1.52E-09	-8.038
53	CG46301	1.77E-09	-1.953
54	CG15483	2.13E-09	-8.447
55	lncRNA:CR45127	2.71E-09	-8.419
56	niki	2.71E-09	-2.644
57	Ugt302K1	2.89E-09	-6.504
58	CG10764	3.43E-09	-3.514
59	Scgalpha	3.55E-09	-1.789
60	mGluR	5.82E-09	-1.723
61	lncRNA:CR45226	6.45E-09	-6.558
62	lncRNA:CR46064	6.68E-09	-1.187
63	CaMKI	6.98E-09	-1.420
64	CG32006	9.16E-09	-1.636
65	CG16965	1.08E-08	-2.699
66	CG11319	1.21E-08	-1.414
67	CG11034	1.42E-08	-1.975
68	CG13305	1.43E-08	-1.676
69	asRNA:CR43730	1.70E-08	-1.934
70	dsb	1.72E-08	-2.836
71	lncRNA:CR44024	2.22E-08	-2.018
72	CG16799	2.41E-08	-1.646
73	CG8768	2.79E-08	-1.592
74	asRNA:CR44031	3.16E-08	-2.023
75	ttm2	3.46E-08	-3.302
76	Had2	5.53E-08	-2.580
77	CG32448	8.13E-08	-1.358
78	asRNA:CR45999	8.39E-08	-1.618
79	asRNA:CR31514	1.02E-07	-2.741
80	AstC-R2	1.02E-07	-1.737
81	CG11158	1.02E-07	-1.509
82	CG11109	1.05E-07	-0.903
83	CG15177	1.73E-07	-7.858
84	CG5961	1.92E-07	-1.119
85	lncRNA:CR46006	2.12E-07	-1.465
86	CG7791	2.36E-07	-0.998
87	CG33465	2.67E-07	-2.537
88	CG4766	2.81E-07	-1.694
89	Arl4	3.05E-07	-0.948
90	DhpD	3.53E-07	-1.366
91	CG1494	3.71E-07	-1.775
92	CG16789	3.92E-07	-1.731
93	lncRNA:CR42719	4.01E-07	-1.992
94	BBS8	4.56E-07	-1.757
95	CG43222	4.58E-07	-1.719
96	CG3216	4.85E-07	-2.427
97	Amt	4.96E-07	-3.810
98	dpr20	5.11E-07	-1.359
99	CR43671	5.23E-07	-4.541
100	asRNA:CR44018	5.47E-07	-2.328

Supplementary Table 2.3: Top 100 downregulated genes with the lowest padj values (T12)

Rank	Gene Symbol	padj value	log2 fc
1	Hsc70-2	1.41E-58	-5.053
2	mthl8	1.95E-55	-6.770
3	w	1.24E-48	-4.809
4	asRNA:CR45822	2.21E-37	-8.646
5	Grik	4.04E-30	-5.009
6	lncRNA:CR43459	4.88E-25	-2.444
7	asRNA:CR44029	1.20E-24	-3.974
8	CG11319	3.54E-23	-2.310
9	CG10407	3.34E-22	-2.611
10	CG9360	4.76E-22	-4.846
11	asRNA:CR45140	1.48E-19	-4.254
12	asRNA:CR45835	2.70E-19	-4.427
13	CG11236	5.70E-18	-4.081
14	SecCl	9.61E-17	-2.003
15	ZnT77C	1.35E-16	-1.941
16	lncRNA:CR44466	9.31E-16	-3.020
17	Ect3	1.15E-15	-2.246
18	CG33337	3.99E-15	-6.973
19	CG31676	7.35E-15	-2.059
20	CG14204	1.15E-14	-4.704
21	lncRNA:CR44743	4.78E-14	-4.355
22	CG14990	5.67E-14	-2.370
23	CG31157	7.55E-14	-3.764
24	NLaz	2.27E-13	-2.878
25	Scgalpha	4.96E-13	-2.155
26	Had2	1.08E-12	-3.408
27	alpha-Est2	2.55E-12	-2.108
28	lncRNA:CR45242	2.86E-12	-3.278
29	lncRNA:CR44754	3.03E-12	-3.603
30	FoxP	4.13E-12	-2.302
31	CG33288	5.19E-12	-2.131
32	AstC-R2	8.04E-12	-2.226
33	Cyp4p2	1.18E-11	-2.359
34	CG14301	1.18E-11	-2.284
35	CG8560	1.75E-11	-2.512
36	NijC	2.30E-11	-1.757
37	Toll-9	3.00E-11	-2.969
38	CG8768	3.42E-11	-1.855
39	CG4766	3.71E-11	-2.185
40	Argk	4.32E-11	-1.882
41	CG33128	8.08E-11	-4.607
42	lncRNA:CR46006	9.10E-11	-1.788
43	CG3842	1.05E-10	-1.705
44	CG42337	1.81E-10	-1.489
45	CG4839	3.19E-10	-1.518
46	CG14694	3.30E-10	-3.792
47	CG3734	3.31E-10	-2.176
48	mGluR	4.73E-10	-1.838
49	Jheh3	5.91E-10	-1.751
50	CG46301	5.94E-10	-2.000

Rank	Gene Symbol	padj value	log2 fc
51	pigs	5.94E-10	-1.709
52	niki	5.95E-10	-3.362
53	dpr21	6.22E-10	-2.377
54	CG32815	7.21E-10	-1.623
55	asRNA:CR44053	1.12E-09	-2.370
56	Ugt302K1	1.39E-09	-5.800
57	CG12502	1.57E-09	-3.076
58	Ugt303B3	2.11E-09	-2.053
59	Cib2	2.60E-09	-1.709
60	Ugt303A1	2.82E-09	-2.113
61	CG10232	2.82E-09	-2.079
62	nAChRalpha7	2.87E-09	-1.728
63	CG1494	3.85E-09	-2.109
64	Spn77Bc	4.34E-09	-3.078
65	ttm2	4.71E-09	-3.984
66	Cyp12c1	5.14E-09	-2.416
67	asRNA:CR31514	5.34E-09	-3.301
68	Cyp6d4	5.66E-09	-2.249
69	lncRNA:CR44603	5.87E-09	-3.298
70	Cyp4ad1	6.09E-09	-1.808
71	CR40190	6.47E-09	-1.114
72	CR42490	8.02E-09	-7.956
73	CG7607	9.42E-09	-2.532
74	CG16799	1.19E-08	-1.751
75	CG4465	1.28E-08	-3.821
76	hiro	1.35E-08	-2.071
77	ATPsynCF6L	1.43E-08	-7.516
78	CG15765	1.67E-08	-1.546
79	Cpr76Bc	1.69E-08	-2.214
80	Ugt49C1	1.74E-08	-2.271
81	lncRNA:CR42719	2.67E-08	-2.408
82	CG43968	2.74E-08	-3.393
83	BBS8	2.81E-08	-1.933
84	CG33465	3.46E-08	-2.832
85	Trh	3.88E-08	-2.309
86	dimm	3.88E-08	-1.799
87	lncRNA:CR44498	3.93E-08	-3.878
88	tap	4.33E-08	-1.996
89	CG5961	4.44E-08	-1.174
90	CG10337	4.58E-08	-1.676
91	mthl4	6.02E-08	-1.819
92	lncRNA:CR46075	6.66E-08	-3.385
93	VhaM9.7-a	6.66E-08	-1.286
94	asRNA:CR45999	6.69E-08	-1.619
95	Teh3	6.99E-08	-1.942
96	lncRNA:CR44467	7.10E-08	-2.059
97	CG6125	8.49E-08	-3.042
98	smp-30	8.54E-08	-2.092
99	Nmdar1	8.54E-08	-1.351
100	lncRNA:CR46064	9.09E-08	-1.124

Supplementary Table 2.4: Top 100 downregulated genes with the lowest padj values (T13)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	3.00E-45	-7.019
2	Hsc70-2	9.29E-41	-3.913
3	w	1.22E-35	-4.060
4	TwdlU	2.77E-32	-5.023
5	CG11236	9.30E-26	-4.210
6	CG10407	8.84E-21	-2.584
7	lncRNA:CR43459	3.91E-19	-2.131
8	asRNA:CR44029	1.42E-18	-3.435
9	CG11319	3.30E-17	-1.894
10	Cyp6a18	6.05E-17	-3.002
11	NLaz	7.10E-16	-3.141
12	CG33337	6.02E-14	-5.668
13	asRNA:CR45835	1.17E-13	-3.317
14	CG6163	1.67E-13	-3.402
15	asRNA:CR45140	3.79E-13	-3.624
16	Nop17l	7.36E-12	-1.054
17	asRNA:CR45822	3.75E-11	-9.048
18	CG33128	4.53E-11	-4.663
19	Cpr76Bc	1.04E-10	-2.415
20	CG31676	1.63E-10	-1.689
21	CG9360	2.67E-10	-3.294
22	CG31157	3.24E-10	-3.338
23	alpha-Est2	4.62E-10	-1.749
24	lncRNA:CR44754	1.30E-09	-2.486
25	Had2	1.50E-09	-2.882
26	CG3842	2.80E-09	-1.609
27	CG10960	3.41E-09	-1.086
28	Grik	4.56E-09	-2.613
29	mthl4	8.94E-09	-1.991
30	CR42490	1.94E-08	-7.927
31	CR40190	1.94E-08	-1.117
32	Syp	2.29E-08	-2.491
33	Cib2	2.61E-08	-1.494
34	CG7408	3.48E-08	-1.753
35	yellow-d2	3.77E-08	-2.721
36	Cyp12e1	4.88E-08	-2.440
37	lncRNA:CR44603	4.93E-08	-2.915
38	ZnT77C	5.67E-08	-1.348
39	Sid	6.58E-08	-2.123
40	CG14990	1.03E-07	-1.728
41	lncRNA:CR44466	1.24E-07	-1.992
42	CG1494	2.35E-07	-1.857
43	lncRNA:CR46006	2.35E-07	-1.499
44	Ugt36E1	3.86E-07	-1.774
45	asRNA:CR44053	3.89E-07	-2.471
46	SecCl	4.61E-07	-1.278
47	Spn77Bc	5.33E-07	-2.698
48	Scgalpha	5.39E-07	-1.574
49	CG11034	6.18E-07	-1.794
50	CaMKI	6.40E-07	-1.268
51	ATPsynCF6L	6.60E-07	-6.837
52	lncRNA:CR45242	8.48E-07	-2.518
53	FoxP	9.07E-07	-1.645
54	Burs	9.24E-07	-1.547
55	Dhfr	9.31E-07	-1.619
56	Dlic	1.04E-06	-0.673
57	CG7607	1.11E-06	-2.713
58	CG14694	1.29E-06	-2.665
59	CG14715	1.34E-06	-2.152
60	Marc	2.10E-06	-1.972
61	CG8560	2.10E-06	-1.855
62	Ect3	2.44E-06	-1.426
63	asRNA:CR44018	2.60E-06	-2.166
64	niki	3.11E-06	-2.082
65	su(w[a])	3.11E-06	-0.672
66	CG9766	3.42E-06	-1.899
67	Cyp311a1	3.82E-06	-2.485
68	Tret1-2	4.50E-06	-2.422
69	CG14406	4.79E-06	-3.745
70	CG6125	5.32E-06	-2.727
71	CG6638	5.60E-06	-1.046
72	Jheh3	7.06E-06	-1.353
73	CG31029	7.44E-06	-2.444
74	CG11997	9.53E-06	-3.226
75	DhpD	9.58E-06	-1.243
76	CG6984	1.01E-05	-1.222
77	CG5961	1.27E-05	-0.999
78	hpRNA:CR33940	1.39E-05	-2.077
79	CG15483	1.53E-05	-6.274
80	CG5599	1.60E-05	-0.977
81	CG8768	1.68E-05	-1.299
82	CG32581	1.92E-05	-0.812
83	CG14456	1.93E-05	-1.574
84	Hmgcl	1.98E-05	-1.231
85	CG4839	2.25E-05	-1.046
86	Trh	2.29E-05	-1.508
87	asRNA:CR45281	2.52E-05	-1.508
88	asRNA:CR31514	2.83E-05	-1.834
89	dpr21	3.26E-05	-1.388
90	ttm2	3.81E-05	-2.879
91	Tina-1	4.54E-05	-0.792
92	CG34195	4.54E-05	-1.149
93	CG3734	4.63E-05	-1.506
94	CG10932	5.40E-05	-1.277
95	CG5171	5.43E-05	-1.812
96	CG33288	5.68E-05	-1.265
97	Ugt303A1	5.89E-05	-1.472
98	CG15553	6.10E-05	-4.241
99	lncRNA:CR44498	6.54E-05	-3.602
100	JhI-26	7.14E-05	-1.480

Supplementary Table 2.5: Top 100 downregulated genes with the lowest padj values (T14)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.82E-60	-6.716
2	w	9.75E-44	-4.637
3	Hsc70-2	2.67E-42	-3.921
4	TwdlU	6.94E-40	-5.793
5	lncRNA:CR43459	1.03E-32	-2.766
6	NLaz	4.58E-31	-4.365
7	CG11236	1.43E-30	-3.283
8	yellow-d2	6.57E-25	-4.974
9	asRNA:CR44029	5.31E-22	-3.715
10	CR42490	9.69E-22	-6.943
11	CG10407	3.49E-19	-2.451
12	CG1494	1.43E-18	-2.965
13	Cyp12e1	1.95E-18	-3.423
14	CG11319	1.94E-17	-1.867
15	Cyp6a18	9.78E-17	-2.540
16	lncRNA:CR44466	1.91E-16	-3.358
17	Syp	3.22E-16	-3.536
18	Ect3	6.74E-16	-2.259
19	CG14406	4.28E-15	-3.237
20	CG42365	1.67E-14	-2.886
21	asRNA:CR42547	6.19E-14	-3.249
22	asRNA:CR45835	6.65E-14	-5.278
23	Nop17l	1.12E-13	-1.123
24	CG31157	1.15E-13	-4.026
25	CG9360	2.73E-13	-3.714
26	Cpr76Bc	3.13E-13	-2.908
27	Grik	3.62E-13	-3.149
28	SecCl	5.44E-13	-1.752
29	CG14456	6.46E-13	-2.082
30	Hmgcl	6.69E-13	-1.928
31	Sid	7.06E-13	-2.695
32	CG13907	8.21E-13	-1.298
33	lncRNA:CR46006	8.70E-13	-1.999
34	CG5171	1.13E-12	-2.900
35	CG31676	1.13E-12	-1.817
36	asRNA:CR45822	3.28E-12	-8.954
37	asRNA:CR45140	4.90E-12	-3.087
38	CG3842	5.51E-12	-1.829
39	CG13813	5.68E-12	-4.232
40	Ugt35B1	1.15E-11	-1.962
41	CG33337	1.16E-11	-3.781
42	Act87E	1.83E-11	-1.336
43	Ir76b	3.02E-11	-2.064
44	CG32369	3.62E-11	-1.366
45	alpha-Est2	4.13E-11	-1.789
46	CG14204	1.04E-10	-4.008
47	CG5961	1.32E-10	-1.379
48	CG14356	1.95E-10	-2.539
49	VhaM9.7-a	3.09E-10	-1.475
50	FoxP	6.29E-10	-1.991

Rank	Gene Symbol	padj value	log2 fc
51	lncRNA:CR44754	8.09E-10	-2.209
52	CG14823	8.63E-10	-3.247
53	fusl	8.93E-10	-1.806
54	Syx4	1.31E-09	-1.472
55	CG6638	1.89E-09	-1.310
56	Invadolysin	3.47E-09	-0.918
57	cmpy	3.83E-09	-1.408
58	CG9801	3.95E-09	-1.393
59	Had2	3.98E-09	-2.750
60	niki	5.81E-09	-2.467
61	Gpb5	6.20E-09	-2.468
62	su(w[a])	6.24E-09	-0.800
63	Scgalpha	7.80E-09	-1.755
64	Ldsdh1	9.30E-09	-1.722
65	CG12099	9.30E-09	-0.940
66	asRNA:CR31514	1.11E-08	-2.424
67	St3	1.14E-08	-1.899
68	Lip1	1.14E-08	-1.969
69	Obp46a	1.57E-08	-3.004
70	CaMKI	1.98E-08	-1.389
71	CG6163	2.04E-08	-2.585
72	lncRNA:CR44024	2.06E-08	-2.057
73	sage	2.12E-08	-1.651
74	CG4250	2.41E-08	-2.017
75	CG4839	2.72E-08	-1.283
76	kar	2.74E-08	-1.579
77	Cib2	3.22E-08	-1.410
78	Mct1	3.71E-08	-1.500
79	Obp99c	3.99E-08	-2.192
80	CG12868	4.03E-08	-1.802
81	lncRNA:CR31386	4.33E-08	-2.720
82	CG8768	6.56E-08	-1.555
83	lncRNA:CR45242	7.80E-08	-2.712
84	lncRNA:CR42719	8.35E-08	-1.999
85	yellow-h	9.87E-08	-1.536
86	CG14990	1.17E-07	-1.693
87	CG32006	1.17E-07	-1.536
88	lncRNA:CR45556	1.18E-07	-3.697
89	Gadd45	1.20E-07	-2.188
90	lncRNA:CR44230	1.27E-07	-3.494
91	Obp99a	1.35E-07	-2.241
92	CG31810	1.98E-07	-1.796
93	CG33124	2.30E-07	-1.985
94	mGluR	2.61E-07	-1.529
95	et	2.79E-07	-1.703
96	CG9766	2.80E-07	-1.928
97	side	3.31E-07	-1.311
98	lncRNA:CR44467	3.65E-07	-1.992
99	CG6044	3.65E-07	-1.095
100	CG13502	3.67E-07	-1.594

Supplementary Table 2.6: Top 100 downregulated genes with the lowest padj values (T15)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	3.14E-62	-7.362
2	w	4.31E-37	-4.152
3	NLaz	8.91E-28	-4.137
4	lncRNA:CR43459	8.91E-28	-2.559
5	Hsc70-2	4.40E-27	-2.961
6	CG11236	1.07E-23	-2.665
7	yellow-d2	3.50E-23	-4.323
8	CG6163	6.24E-22	-4.434
9	CG1494	7.70E-22	-3.147
10	Cyp12e1	2.11E-21	-3.558
11	asRNA:CR44029	5.52E-21	-3.577
12	CG14406	2.30E-18	-2.918
13	asRNA:CR45835	7.48E-17	-4.280
14	TwdlU	2.60E-16	-4.116
15	yip2	4.21E-16	-1.959
16	asRNA:CR45822	3.33E-15	-8.007
17	CG31157	8.36E-15	-4.240
18	CG10407	2.48E-14	-2.103
19	Grik	2.87E-14	-3.254
20	Hmgcl	7.72E-14	-2.001
21	Cyp6a18	9.77E-14	-2.210
22	Nop17l	1.04E-13	-1.119
23	Ect3	1.45E-12	-2.004
24	CG14694	1.05E-11	-3.269
25	hpRNA:CR33940	1.21E-11	-2.956
26	CG14204	1.26E-11	-4.606
27	CR40190	1.54E-11	-1.290
28	CG31676	1.64E-11	-1.716
29	lncRNA:CR44466	2.22E-11	-2.732
30	Syp	2.99E-11	-3.072
31	CG11319	4.12E-11	-1.467
32	asRNA:CR45140	4.36E-11	-2.855
33	CG34195	1.59E-10	-1.735
34	CG13813	1.73E-10	-3.953
35	CG33128	2.27E-10	-4.145
36	Lip1	3.73E-10	-2.132
37	CG14715	3.82E-10	-2.663
38	CG5599	3.82E-10	-1.312
39	lncRNA:CR46006	4.15E-10	-1.742
40	SecCl	5.36E-10	-1.506
41	CG5938	1.06E-09	-0.846
42	CG13827	1.12E-09	-1.533
43	CG7402	1.37E-09	-2.514
44	CG5171	1.99E-09	-2.487
45	CR42490	2.76E-09	-8.060
46	CG14823	3.13E-09	-2.881
47	Obp99a	3.60E-09	-2.462
48	CG32581	3.60E-09	-1.045
49	CG34276	3.97E-09	-2.183
50	CG6984	4.54E-09	-1.520
51	Obp46a	4.86E-09	-3.256
52	CG14456	6.81E-09	-1.547
53	Snx21	1.03E-08	-1.199
54	Ugt35B1	1.51E-08	-1.534
55	CG13445	1.56E-08	-4.122
56	FoxP	1.97E-08	-1.776
57	CG6638	2.01E-08	-1.226
58	CG5961	2.36E-08	-1.203
59	CG13907	2.87E-08	-1.032
60	CG42365	4.45E-08	-2.127
61	Spn77Bc	4.59E-08	-2.784
62	su(w[a])	4.68E-08	-0.754
63	CG10232	8.18E-08	-1.789
64	CG12896	8.92E-08	-2.686
65	CG6431	1.05E-07	-4.250
66	lncRNA:CR31386	1.20E-07	-2.165
67	CaMKI	1.38E-07	-1.308
68	Mccc2	1.89E-07	-1.337
69	amd	1.97E-07	-2.498
70	kar	2.00E-07	-1.486
71	dpr21	2.00E-07	-1.421
72	alpha-Est2	2.12E-07	-1.429
73	lncRNA:CR45556	2.19E-07	-4.478
74	Pgant3	2.81E-07	-1.464
75	Act87E	3.05E-07	-1.051
76	Invadolysin	3.09E-07	-0.809
77	ATPsynCF6L	3.27E-07	-6.813
78	CG4766	3.42E-07	-1.766
79	CG43968	3.55E-07	-2.481
80	CG6836	4.06E-07	-1.965
81	ect	5.74E-07	-3.230
82	St3	6.42E-07	-1.679
83	asRNA:CR44179	7.13E-07	-1.502
84	niki	7.14E-07	-1.999
85	lncRNA:CR44754	7.14E-07	-1.585
86	CG13312	8.63E-07	-3.691
87	asRNA:CR44018	9.88E-07	-2.046
88	Ir76b	9.99E-07	-1.468
89	Orct	1.12E-06	-1.693
90	Prx2540-1	1.50E-06	-2.479
91	CG7255	1.89E-06	-1.034
92	SCOT	1.92E-06	-0.912
93	CG9766	1.95E-06	-1.703
94	CG31029	2.83E-06	-2.390
95	Gpb5	2.87E-06	-1.825
96	Sid	2.91E-06	-1.828
97	Syt12	3.09E-06	-1.162
98	Etf-QO	3.68E-06	-0.857
99	Cib2	3.96E-06	-1.166
100	CR41379	4.17E-06	-1.843

Supplementary Table 2.7: Top 100 downregulated genes with the lowest padj values (T16)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	6.77E-73	-6.334
2	CG14406	1.21E-51	-5.214
3	NLaz	9.70E-38	-4.814
4	w	1.26E-34	-4.111
5	CG1494	3.89E-34	-3.971
6	CG11236	1.66E-30	-3.128
7	yellow-d2	4.27E-30	-5.293
8	Hsc70-2	1.30E-28	-3.164
9	Ect3	6.90E-27	-2.954
10	CG11319	6.02E-26	-2.249
11	Grik	3.75E-25	-4.495
12	CG14661	4.81E-25	-3.295
13	CG42807	1.87E-23	-3.159
14	lncRNA:CR43459	2.02E-23	-2.361
15	CG9864	8.53E-23	-3.256
16	CG3690	8.63E-22	-3.520
17	asRNA:CR44029	1.14E-21	-3.694
18	CG6296	2.68E-21	-3.182
19	CG13159	4.47E-20	-5.625
20	Ugt35B1	4.50E-20	-2.457
21	CG3819	7.50E-20	-5.152
22	Cyp12e1	8.70E-20	-3.418
23	lncRNA:CR44754	9.15E-19	-2.808
24	Cyp6a18	9.15E-19	-2.597
25	CG13813	1.05E-18	-5.335
26	asRNA:CR45835	1.22E-18	-5.646
27	MFS14	2.96E-18	-1.893
28	CG10407	6.82E-18	-2.323
29	CG14694	1.53E-17	-4.796
30	CG14456	1.53E-17	-2.137
31	CR42490	2.30E-16	-6.495
32	dpr21	3.39E-16	-2.111
33	Btnd	7.09E-16	-3.184
34	CG31676	7.59E-16	-2.036
35	Act87E	1.93E-15	-1.549
36	lncRNA:CR46006	1.98E-15	-2.190
37	CG10232	4.87E-15	-2.513
38	CG1304	5.64E-15	-5.135
39	CG11413	6.23E-15	-3.229
40	CG11000	1.07E-14	-1.682
41	SecCl	1.34E-14	-1.804
42	Ir76b	1.40E-14	-2.231
43	CG14823	2.37E-14	-3.826
44	Syt12	2.58E-14	-1.787
45	yip2	4.40E-14	-1.824
46	Lk	4.71E-14	-3.938
47	CG34276	1.57E-13	-2.777
48	CG5171	1.58E-13	-2.997
49	Ugt35E1	1.79E-13	-2.488
50	NaPi-T	2.08E-13	-2.765
51	Nmdar1	3.61E-13	-1.688
52	CG5961	3.63E-13	-1.511
53	CG7300	3.66E-13	-3.372
54	CG32815	5.66E-13	-1.835
55	Obp56b	6.13E-13	-3.424
56	asRNA:CR42547	8.26E-13	-2.837
57	btl	1.94E-12	-2.348
58	CG3270	2.27E-12	-2.874
59	Obp99a	3.06E-12	-2.854
60	RunxA	3.24E-12	-2.467
61	alpha-Est2	4.16E-12	-1.837
62	CG8664	4.55E-12	-3.009
63	Oseg4	4.78E-12	-2.044
64	kar	4.81E-12	-1.924
65	CG30026	5.37E-12	-1.943
66	CG31157	8.76E-12	-3.643
67	CG30091	8.99E-12	-3.489
68	SiaT	1.05E-11	-2.405
69	CG5402	1.38E-11	-2.679
70	CG10877	1.53E-11	-1.272
71	lncRNA:CR31386	2.18E-11	-2.909
72	CG6847	3.03E-11	-2.114
73	CG8003	3.09E-11	-1.542
74	CG16854	4.17E-11	-1.807
75	CG18095	5.16E-11	-1.931
76	CG32581	5.16E-11	-1.141
77	CG43968	5.81E-11	-2.956
78	Spn77Bc	5.87E-11	-3.373
79	pigs	6.00E-11	-1.779
80	CG14204	6.34E-11	-3.818
81	Orct	7.46E-11	-2.235
82	CG14234	8.52E-11	-2.132
83	LysD	1.00E-10	-3.694
84	lncRNA:CR31781	1.06E-10	-1.786
85	CR40190	1.33E-10	-1.219
86	SCOT	2.09E-10	-1.164
87	CG12896	2.25E-10	-3.105
88	CG42540	2.51E-10	-1.648
89	asRNA:CR44179	2.53E-10	-1.863
90	CG4766	2.87E-10	-2.265
91	CG11905	2.96E-10	-2.209
92	Gpb5	3.32E-10	-2.340
93	VhaM9.7-a	3.79E-10	-1.450
94	snu	5.01E-10	-1.645
95	CG5246	5.19E-10	-3.206
96	Pxd	5.90E-10	-1.490
97	CG9826	7.03E-10	-2.483
98	natalisin	7.38E-10	-2.022
99	asRNA:CR45822	8.40E-10	-8.161
100	Orcokinin	8.40E-10	-2.073

Supplementary Table 2.8: Top 100 downregulated genes with the lowest padj values (T17)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	8.84E-61	-5.974
2	CG14406	1.66E-51	-5.532
3	CG1494	1.97E-36	-4.178
4	w	9.15E-34	-4.408
5	asRNA:CR44179	7.99E-32	-3.352
6	lncRNA:CR43459	5.23E-31	-2.738
7	yellow-d2	6.63E-25	-5.902
8	NLaz	6.63E-25	-3.936
9	CG11319	6.66E-25	-2.208
10	CG11236	4.65E-24	-2.821
11	Hsc70-2	2.29E-23	-2.878
12	lncRNA:CR44754	1.10E-21	-3.276
13	CG14961	1.99E-21	-4.058
14	Cyp6a18	1.18E-20	-2.834
15	asRNA:CR44029	5.47E-20	-3.570
16	Grik	1.58E-19	-4.049
17	Ect3	1.78E-19	-2.535
18	CG42807	2.32E-19	-2.593
19	CG14661	4.04E-19	-2.744
20	CG11000	2.33E-18	-1.891
21	CG30091	5.08E-18	-4.205
22	CG6296	1.71E-17	-2.863
23	CG10407	1.98E-17	-2.302
24	dpr21	2.55E-17	-2.315
25	Syt12	2.58E-17	-2.003
26	CG9864	1.00E-16	-2.781
27	asRNA:CR45822	4.46E-16	-6.959
28	CG3690	6.19E-16	-3.069
29	CG13813	1.27E-15	-4.985
30	CG1304	1.44E-15	-5.253
31	Lk	1.49E-15	-4.166
32	CG33128	1.68E-15	-5.267
33	CG5961	1.94E-15	-1.653
34	CG31676	2.87E-15	-2.016
35	CG32815	3.19E-15	-2.013
36	asRNA:CR45835	6.31E-15	-5.431
37	Cyp12e1	9.26E-15	-2.960
38	Nmdar1	1.78E-14	-1.779
39	Nop17l	1.78E-14	-1.151
40	Obp56b	2.02E-14	-3.881
41	CG14694	2.53E-14	-3.609
42	Ir76b	3.02E-14	-2.214
43	pigs	3.49E-14	-2.042
44	CG3819	4.85E-14	-4.335
45	Ugt35E1	7.08E-14	-2.536
46	lncRNA:CR31781	1.46E-13	-2.027
47	side	2.43E-13	-1.799
48	Btnd	2.59E-13	-2.921
49	lncRNA:CR46006	4.32E-13	-2.036
50	ZnT77C	6.38E-13	-1.687
51	bma	1.04E-12	-1.718
52	CG3270	1.09E-12	-2.911
53	NaPi-T	1.97E-12	-2.664
54	Orcokinin	2.11E-12	-2.417
55	Cyp4ac3	2.47E-12	-3.043
56	CG10232	3.81E-12	-2.266
57	Oseg4	8.03E-12	-2.034
58	MFS14	8.23E-12	-1.525
59	CG6163	8.40E-12	-3.046
60	SecCl	8.78E-12	-1.619
61	Hug	9.10E-12	-3.111
62	CG14234	1.17E-11	-2.232
63	RunxA	1.22E-11	-2.529
64	GlyT	1.22E-11	-2.175
65	CG14456	1.86E-11	-1.711
66	CG13743	2.00E-11	-1.533
67	asRNA:CR42547	2.05E-11	-2.876
68	CG33337	2.11E-11	-3.007
69	CG42540	2.31E-11	-1.738
70	Ugt35B1	2.60E-11	-1.905
71	Capa	2.77E-11	-3.592
72	Pxd	2.88E-11	-1.608
73	natalisin	3.36E-11	-2.186
74	side-III	3.48E-11	-1.493
75	Tomosyn	3.74E-11	-1.440
76	Hasp	4.10E-11	-2.031
77	lncRNA:CR42646	5.34E-11	-3.945
78	Appl	5.35E-11	-1.616
79	CG8003	5.52E-11	-1.531
80	lncRNA:CR31386	5.89E-11	-3.431
81	CG6218	5.89E-11	-1.592
82	CG6688	7.24E-11	-2.021
83	nSyb	7.69E-11	-1.522
84	CG31191	7.94E-11	-1.396
85	CG34354	9.43E-11	-1.732
86	Dscam2	1.14E-10	-1.414
87	Ac13E	1.28E-10	-0.953
88	CG30026	1.42E-10	-1.865
89	CG1090	1.74E-10	-1.601
90	tbc	1.93E-10	-1.958
91	CG14880	2.16E-10	-1.762
92	CG12896	2.62E-10	-3.105
93	CG6638	2.63E-10	-1.361
94	CG16854	2.77E-10	-1.785
95	jus	2.77E-10	-1.552
96	CG31157	3.24E-10	-3.460
97	beat-Vc	3.61E-10	-1.634
98	Lip1	4.75E-10	-2.185
99	yip2	5.05E-10	-1.534
100	CG4577	5.39E-10	-2.039

Supplementary Table 2.9: Top 100 downregulated genes with the lowest padj values (T18)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	2.66E-50	-6.129
2	CG14406	2.77E-49	-6.271
3	CG42807	3.27E-42	-3.253
4	CG3690	3.97E-38	-3.665
5	lncRNA:CR43459	7.70E-38	-3.024
6	CG6296	6.77E-37	-4.149
7	w	1.44E-33	-4.518
8	CG1494	2.43E-32	-3.896
9	CG11236	3.00E-32	-3.340
10	CG3819	3.16E-30	-5.478
11	NaPi-T	2.84E-29	-3.939
12	CG11319	2.38E-28	-2.340
13	asRNA:CR45822	3.96E-28	-6.521
14	Btnd	3.28E-26	-3.922
15	Ugt35E1	5.30E-26	-3.347
16	Ugt35B1	1.45E-25	-3.129
17	lncRNA:CR44754	3.99E-25	-3.800
18	dpr21	4.57E-24	-2.574
19	AcCoAS	1.91E-23	-2.102
20	Hsc70-2	3.84E-21	-2.807
21	Cyp6a18	4.10E-21	-2.885
22	CG11000	4.37E-21	-2.007
23	Grik	6.43E-21	-4.326
24	CG5961	6.43E-21	-1.908
25	asRNA:CR45835	1.09E-20	-4.422
26	Syt12	1.80E-20	-2.172
27	Obp99d	2.20E-20	-3.567
28	CG1304	4.86E-20	-5.925
29	asRNA:CR44029	2.14E-19	-3.568
30	CG14661	1.33E-18	-2.620
31	ZnT77C	3.38E-18	-2.012
32	Mal-A8	3.66E-18	-3.053
33	Pxd	4.46E-18	-2.104
34	lncRNA:CR31781	9.44E-18	-2.315
35	CG6733	1.56E-17	-3.337
36	CG33337	1.64E-17	-3.663
37	CG30026	2.85E-17	-2.412
38	MFS14	3.01E-17	-1.833
39	CG3270	3.19E-17	-3.350
40	beta4GalNAcTA	5.22E-17	-1.670
41	lncRNA:CR46006	1.34E-16	-2.303
42	CG9826	1.56E-16	-3.086
43	CG18327	1.70E-16	-2.971
44	CG31676	1.73E-16	-2.083
45	CG5246	2.10E-16	-3.803
46	CG8745	4.01E-16	-3.120
47	Gpb5	1.51E-15	-2.961
48	alpha-Est2	4.34E-15	-2.050
49	CR40190	4.34E-15	-1.484
50	SecCl	4.35E-15	-1.796
51	Orcokinin	4.54E-15	-2.524
52	DIP-lambda	4.84E-15	-1.938
53	CG9864	5.44E-15	-2.600
54	yellow-d2	5.71E-15	-5.588
55	Obp56b	1.07E-14	-3.466
56	CG32581	1.12E-14	-1.324
57	CG10407	1.20E-14	-2.066
58	SCOT	1.21E-14	-1.380
59	Cyp12e1	1.23E-14	-2.980
60	nAChRalpha7	1.25E-14	-1.932
61	Ect3	1.92E-14	-2.165
62	Ir76b	2.36E-14	-2.232
63	Gal	2.66E-14	-2.195
64	CG10361	2.83E-14	-2.367
65	CG31191	3.35E-14	-1.597
66	CG32302	4.47E-14	-2.939
67	NijC	4.63E-14	-1.963
68	daw	5.06E-14	-2.050
69	Act87E	5.60E-14	-1.463
70	Mccc2	7.08E-14	-1.821
71	CG5599	7.08E-14	-1.513
72	CG14823	7.32E-14	-4.453
73	Vha26	1.07E-13	-1.207
74	Vha16-1	1.07E-13	-1.164
75	CG42540	1.15E-13	-1.888
76	CG30091	1.19E-13	-3.524
77	CG18095	1.29E-13	-2.144
78	CG3902	1.45E-13	-1.807
79	CG33124	1.69E-13	-2.623
80	CG43778	1.74E-13	-1.899
81	Ggamma30A	1.76E-13	-1.802
82	CG2765	1.97E-13	-0.868
83	Nmdar1	1.98E-13	-1.694
84	CG31446	1.99E-13	-2.130
85	bma	3.07E-13	-1.732
86	CG32379	3.15E-13	-4.855
87	CG31097	4.86E-13	-2.167
88	Dh44	5.00E-13	-1.729
89	CG4752	5.02E-13	-1.596
90	CG8199	5.03E-13	-1.685
91	CG8664	5.13E-13	-2.832
92	Gba1a	5.41E-13	-2.500
93	DAT	6.22E-13	-2.107
94	Nop17l	6.38E-13	-1.074
95	NLaz	6.63E-13	-2.791
96	FASN1	7.66E-13	-2.307
97	lncRNA:CR42719	9.68E-13	-2.431
98	Cyp9h1	1.03E-12	-2.122
99	CG40470	1.54E-12	-1.886
100	Cbp53E	1.64E-12	-1.797

Supplementary Table 2.10: Top 100 downregulated genes with the lowest padj values (T19)

Rank	Gene Symbol	padj value	log2 fc
1	CG3819	1.21E-47	-6.621
2	CG14406	1.03E-41	-6.876
3	mthl8	2.89E-41	-6.721
4	CG3690	1.15E-36	-3.271
5	CG1494	6.61E-35	-4.463
6	lncRNA:CR43459	3.02E-30	-2.721
7	CG11236	7.95E-30	-3.148
8	w	6.57E-28	-3.843
9	CG42807	1.82E-27	-2.519
10	CG5961	2.53E-26	-2.143
11	asRNA:CR45822	2.88E-26	-6.754
12	NaPi-T	4.35E-26	-3.654
13	CG11319	6.23E-25	-2.196
14	Hsc70-2	6.89E-25	-3.035
15	lncRNA:CR44754	2.47E-22	-3.478
16	CG6733	3.38E-22	-3.782
17	CG33337	1.12E-21	-4.040
18	CG6296	8.29E-21	-2.976
19	CG10407	1.79E-20	-2.478
20	Btnd	3.27E-19	-3.311
21	Ugt35E1	4.92E-19	-2.773
22	nop5	5.22E-19	-0.895
23	CG3270	5.83E-19	-3.510
24	CG1304	7.76E-19	-5.757
25	AcCoAS	1.78E-18	-1.869
26	Grik	1.02E-17	-3.859
27	asRNA:CR44029	2.56E-17	-3.331
28	Ugt35B1	3.60E-17	-2.545
29	CG18327	8.69E-17	-2.927
30	asRNA:CR45835	2.10E-16	-5.034
31	MFS14	2.24E-16	-1.795
32	Obp99d	2.64E-16	-3.065
33	Nop17l	7.53E-16	-1.205
34	Art8	8.24E-16	-1.530
35	Mal-A8	2.30E-15	-2.849
36	CG3902	2.42E-15	-1.940
37	CG9826	4.36E-15	-2.909
38	CG5402	4.79E-15	-3.256
39	CG11000	4.79E-15	-1.693
40	Ect3	5.29E-15	-2.239
41	Syt12	1.17E-14	-1.829
42	CG17219	1.65E-14	-1.752
43	CG5246	1.65E-14	-3.238
44	Gpb5	2.26E-14	-3.008
45	CG8745	5.86E-14	-2.926
46	Gal	1.07E-13	-2.157
47	Mccc2	1.42E-13	-1.816
48	dpr21	1.69E-13	-1.882
49	Lk	1.72E-13	-3.495
50	Gba1a	1.72E-13	-2.540
51	CG9799	2.84E-13	-1.575
52	lncRNA:CR44466	3.49E-13	-3.594
53	Orcokinin	3.85E-13	-2.338
54	CG33124	4.80E-13	-2.607
55	lncRNA:CR31781	5.82E-13	-1.982
56	ACC	9.72E-13	-1.795
57	DAT	1.01E-12	-2.120
58	beta4GalNAcTA	1.20E-12	-1.445
59	NHP2	1.20E-12	-1.274
60	FASN1	1.24E-12	-2.310
61	Cyp6a18	1.24E-12	-2.204
62	Uhg2	1.60E-12	-1.823
63	CG14823	2.30E-12	-3.859
64	CG1671	2.34E-12	-1.519
65	Gclc	2.50E-12	-1.357
66	l(1)G0020	3.65E-12	-1.543
67	CG7966	3.91E-12	-1.864
68	CG5599	3.91E-12	-1.422
69	CG11158	4.56E-12	-1.928
70	ZnT77C	4.83E-12	-1.633
71	CG6638	4.83E-12	-1.471
72	NijC	7.64E-12	-1.924
73	CG8064	8.80E-12	-1.448
74	Cyp305a1	1.23E-11	-1.706
75	PH4alphaMP	1.23E-11	-2.451
76	lncRNA:CR46006	1.32E-11	-1.895
77	CG7458	1.48E-11	-1.143
78	CG10361	1.56E-11	-2.116
79	CG14298	1.56E-11	-2.037
80	CG16775	1.88E-11	-3.582
81	CG31191	3.08E-11	-1.415
82	CG32302	3.16E-11	-2.624
83	CG8199	4.52E-11	-1.563
84	CG11034	4.90E-11	-2.220
85	Orct	5.68E-11	-2.238
86	Hmgcl	6.23E-11	-1.723
87	CG13907	6.38E-11	-1.192
88	DIP-lambda	7.18E-11	-1.652
89	natalisin	8.91E-11	-2.038
90	yellow-d2	1.18E-10	-3.836
91	Obp56b	1.22E-10	-2.876
92	Mys45A	1.25E-10	-1.081
93	CG31676	1.63E-10	-1.656
94	CG14661	2.08E-10	-1.916
95	CG4783	2.85E-10	-8.661
96	CG31157	2.85E-10	-3.497
97	Pxd	2.87E-10	-1.656
98	CG17026	3.27E-10	-2.226
99	Had2	3.30E-10	-2.558
100	Nmdar1	3.31E-10	-1.480

Supplementary Table 2.11: Top 100 downregulated genes with the lowest padj values (T20)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	5.72E-41	-6.948
2	CG3819	5.16E-40	-5.936
3	lncRNA:CR43459	6.03E-33	-2.857
4	Hsc70-2	1.04E-31	-3.384
5	CG3690	2.90E-27	-2.703
6	Obp99d	4.94E-25	-3.749
7	w	2.52E-24	-3.523
8	CG14406	1.03E-20	-8.079
9	nop5	2.22E-20	-0.931
10	CG34195	5.16E-20	-2.447
11	CG33337	1.94E-19	-3.949
12	asRNA:CR45835	5.68E-18	-4.098
13	Grik	4.44E-17	-3.864
14	CG3270	2.04E-16	-3.294
15	Ugt35B1	2.35E-16	-3.095
16	CG5961	6.63E-16	-1.670
17	CG11236	7.67E-16	-2.208
18	CG18327	1.16E-15	-2.836
19	CG10474	1.25E-15	-5.445
20	CG1494	1.85E-15	-2.710
21	CG1304	3.74E-15	-5.187
22	CG5402	2.43E-14	-3.335
23	NaPi-T	2.51E-14	-2.698
24	lncRNA:CR44754	2.85E-14	-3.193
25	aay	6.10E-14	-0.933
26	CG6296	8.04E-14	-2.414
27	CG17219	8.97E-14	-1.697
28	CG7458	1.06E-13	-1.264
29	CG6712	1.34E-13	-1.364
30	CG33124	1.47E-13	-2.699
31	CG10407	3.88E-13	-1.973
32	CG11319	3.88E-13	-1.596
33	CG9515	6.11E-13	-1.375
34	U3-55K	6.17E-13	-1.517
35	CG42807	6.83E-13	-1.686
36	Mys45A	6.83E-13	-1.201
37	Art8	8.03E-13	-1.379
38	CG6733	1.18E-12	-2.857
39	NHP2	1.41E-12	-1.274
40	firl	1.77E-12	-2.463
41	CG16775	6.25E-12	-3.840
42	asRNA:CR45822	6.27E-12	-8.667
43	CG11158	1.11E-11	-1.906
44	PH4alphaMP	1.30E-11	-2.505
45	CG6763	2.22E-11	-2.756
46	Fit2	2.52E-11	-1.347
47	CG13003	2.65E-11	-2.104
48	l(2)09851	3.43E-11	-1.056
49	Lk	3.52E-11	-3.210
50	CG8064	4.09E-11	-1.412
51	asRNA:CR44029	6.19E-11	-2.607
52	lncRNA:CR46006	6.64E-11	-1.878
53	CG10175	8.09E-11	-1.962
54	Nop56	1.62E-10	-1.062
55	Cpr51A	1.70E-10	-2.729
56	CG30091	2.65E-10	-3.035
57	CG7227	2.65E-10	-2.438
58	DAT	2.79E-10	-1.907
59	CG6638	2.79E-10	-1.362
60	Capa	3.21E-10	-3.030
61	pit	3.87E-10	-1.089
62	CG15083	4.13E-10	-1.326
63	Ugt35E1	4.29E-10	-2.010
64	Dph3	6.82E-10	-1.149
65	AcCoAS	8.62E-10	-1.356
66	Mtr3	9.16E-10	-1.099
67	Mccc2	1.08E-09	-1.530
68	yellow-d2	1.10E-09	-4.043
69	Pus1	1.14E-09	-1.571
70	Nop17l	1.53E-09	-0.927
71	Mal-A8	1.59E-09	-2.189
72	ChLD3	2.00E-09	-3.810
73	CG14598	2.24E-09	-1.969
74	CG14204	2.91E-09	-3.276
75	CG9799	3.07E-09	-1.308
76	CG7300	3.12E-09	-2.695
77	CG15877	3.27E-09	-1.215
78	CG18273	3.38E-09	-1.012
79	Obp56b	3.45E-09	-2.639
80	CG13606	3.54E-09	-1.847
81	MFS14	3.63E-09	-1.333
82	dpr21	5.16E-09	-1.541
83	CG13907	5.29E-09	-1.083
84	CG2972	5.48E-09	-0.887
85	Aprt	7.03E-09	-1.243
86	Mccc1	8.19E-09	-1.207
87	lncRNA:CR46064	9.59E-09	-1.241
88	Syt12	9.75E-09	-1.399
89	l(1)G0020	1.04E-08	-1.300
90	CG43181	1.06E-08	-4.469
91	alpha-Est2	1.06E-08	-1.585
92	CG34437	1.12E-08	-2.331
93	CG31157	1.12E-08	-3.070
94	CG6724	1.16E-08	-1.259
95	Pxd	1.22E-08	-1.537
96	CG14298	1.49E-08	-1.733
97	CG8745	1.89E-08	-2.259
98	Gba1a	2.51E-08	-1.981
99	CG5599	2.74E-08	-1.169
100	CG1785	3.50E-08	-0.873

Supplementary Table 2.12: Top 100 downregulated genes with the lowest padj values (T21)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.82E-52	-5.952
2	Hsc70-2	1.66E-43	-4.141
3	CG3819	7.77E-42	-6.100
4	CG3690	1.56E-34	-3.058
5	asRNA:CR45822	2.74E-33	-6.185
6	Obp99d	3.10E-29	-4.099
7	lncRNA:CR43459	1.26E-28	-2.648
8	CG14406	5.40E-26	-7.290
9	w	2.11E-23	-3.522
10	Grik	1.61E-20	-4.347
11	CG11319	2.47E-20	-1.989
12	asRNA:CR45835	5.63E-20	-4.483
13	CG42807	1.52E-19	-2.085
14	Nop17l	1.93E-18	-1.298
15	CG1494	9.66E-18	-2.954
16	CG11236	3.13E-16	-2.247
17	aay	9.52E-16	-0.989
18	CG6296	1.22E-15	-2.567
19	Oseg4	2.61E-15	-2.348
20	CG5961	3.46E-15	-1.634
21	CG1304	8.80E-15	-5.127
22	Lk	2.17E-14	-3.824
23	CG13907	1.04E-13	-1.355
24	CG10407	1.20E-13	-2.002
25	dpr21	2.42E-13	-1.929
26	CG3270	5.38E-13	-2.929
27	Gpb5	1.03E-12	-2.951
28	CG33337	1.16E-12	-3.140
29	CG34195	1.93E-12	-1.825
30	Pfas	2.39E-12	-1.323
31	CG5599	3.17E-12	-1.439
32	Mal-A8	1.01E-11	-2.453
33	NaPi-T	1.66E-11	-2.419
34	asRNA:CR44029	2.30E-11	-2.653
35	CG2765	2.51E-11	-0.816
36	Mccc2	3.01E-11	-1.661
37	CG12896	4.98E-11	-3.256
38	CG10474	6.47E-11	-4.491
39	U3-55K	7.40E-11	-1.392
40	CG6638	8.52E-11	-1.405
41	CG11000	9.32E-11	-1.439
42	lncRNA:CR44754	1.07E-10	-2.466
43	CAH2	1.58E-10	-1.316
44	CG18327	2.32E-10	-2.265
45	Obp56b	3.18E-10	-2.820
46	CG10932	3.96E-10	-1.807
47	l(2)09851	7.80E-10	-0.992
48	CG8064	8.34E-10	-1.330
49	Cpr64Ac	1.07E-09	-6.785
50	CG31157	1.07E-09	-3.214

Rank	Gene Symbol	padj value	log2 fc
51	CG13159	1.84E-09	-3.912
52	Dph3	2.31E-09	-1.121
53	CG5402	2.35E-09	-2.595
54	CG11158	2.39E-09	-1.715
55	LysD	2.67E-09	-3.270
56	CG6163	3.65E-09	-2.745
57	nop5	3.71E-09	-0.621
58	Prx2540-1	4.52E-09	-2.972
59	ChLD3	6.11E-09	-3.634
60	CG3902	6.34E-09	-1.478
61	CG8199	7.91E-09	-1.402
62	Mccc1	8.51E-09	-1.212
63	Art8	1.26E-08	-1.126
64	Hmgcl	1.32E-08	-1.529
65	MFS14	1.56E-08	-1.293
66	SelR	1.75E-08	-1.329
67	Pus1	1.90E-08	-1.473
68	Mtr3	1.95E-08	-1.023
69	CG6733	2.02E-08	-2.335
70	CG16727	2.16E-08	-2.120
71	CG8745	2.18E-08	-2.263
72	Capa	3.52E-08	-2.696
73	CG3348	3.92E-08	-3.819
74	ValRS	4.07E-08	-1.056
75	Btnd	4.15E-08	-2.135
76	CG11779	4.34E-08	-1.016
77	Spn77Bc	4.46E-08	-3.129
78	CG4783	4.75E-08	-7.799
79	yellow-d2	4.75E-08	-4.118
80	lncRNA:roX2	4.78E-08	-1.778
81	CG9799	5.61E-08	-1.217
82	l(1)G0020	5.71E-08	-1.247
83	CG34005	6.25E-08	-3.520
84	Syt12	6.29E-08	-1.334
85	blp	6.45E-08	-1.008
86	ZnT77C	1.00E-07	-1.305
87	firl	1.29E-07	-1.871
88	AdSL	1.29E-07	-1.038
89	lncRNA:CR31781	1.30E-07	-1.514
90	CG17219	1.30E-07	-1.220
91	lncRNA:CR45223	1.38E-07	-1.879
92	Mys45A	1.38E-07	-0.913
93	NijC	1.48E-07	-1.580
94	CG6712	1.51E-07	-1.003
95	CG33124	1.59E-07	-1.982
96	Ugt35B1	1.71E-07	-1.831
97	ATP7	1.93E-07	-0.876
98	PH4alphaMP	2.16E-07	-1.976
99	NHP2	2.17E-07	-0.968
100	CG9281	2.44E-07	-0.821

Supplementary Table 2.13: Top 100 downregulated genes with the lowest padj values (T22)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	5.83E-55	-7.742
2	CG3819	3.37E-37	-5.676
3	Grik	4.63E-34	-5.493
4	CG14406	4.17E-31	-8.008
5	CG9360	8.11E-31	-5.666
6	CG10407	1.08E-29	-2.884
7	firl	3.35E-29	-3.741
8	lncRNA:CR43459	2.46E-28	-2.626
9	CG33337	1.19E-27	-4.905
10	Hsc70-2	1.98E-27	-2.979
11	Obp99d	2.89E-27	-4.292
12	GstE1	6.56E-26	-3.279
13	Keap1	2.60E-24	-1.539
14	CG30091	7.34E-23	-4.497
15	CG14186	3.46E-22	-3.301
16	Cpr51A	5.12E-21	-3.889
17	w	5.12E-21	-3.251
18	ZnT77C	1.68E-20	-2.124
19	asRNA:CR44179	6.32E-20	-2.722
20	CG6296	1.99E-18	-2.768
21	Ect3	2.86E-18	-2.468
22	CG9733	6.47E-18	-4.916
23	NLaz	7.56E-18	-3.322
24	asRNA:CR45835	1.12E-17	-5.173
25	CG2150	1.53E-17	-5.286
26	Oseg4	2.91E-17	-2.492
27	Obp99a	6.23E-17	-3.364
28	Cpr64Ac	7.18E-17	-9.004
29	Phk-3	7.18E-17	-2.600
30	Sb	1.08E-16	-4.441
31	CG42807	1.11E-16	-1.926
32	CG13159	1.20E-16	-5.140
33	CG5402	1.72E-16	-3.529
34	CG30026	3.61E-16	-2.349
35	TwdlX	4.24E-16	-2.194
36	CG43204	5.57E-16	-3.724
37	lncRNA:CR44754	5.57E-16	-3.252
38	CG7406	6.89E-16	-6.665
39	CG11413	6.89E-16	-3.276
40	CG31869	1.17E-15	-2.465
41	scaf	1.51E-15	-2.465
42	asRNA:CR45822	2.21E-15	-8.055
43	lncRNA:CR42646	3.25E-15	-4.595
44	CG14567	6.56E-15	-4.585
45	ChLD3	6.68E-15	-4.835
46	CG14961	8.08E-15	-3.266
47	Gadd45	8.91E-15	-3.039
48	CG7724	1.79E-14	-2.068
49	asRNA:CR45140	1.96E-14	-3.224
50	ThrRS	2.32E-14	-0.917

Rank	Gene Symbol	padj value	log2 fc
51	Muc91C	3.09E-14	-6.742
52	CG13606	3.09E-14	-2.376
53	Dyrk2	3.09E-14	-1.271
54	CG34437	4.82E-14	-3.220
55	TwdlC	5.04E-14	-6.512
56	Cpr65Ea	5.04E-14	-4.352
57	Cpr35B	5.04E-14	-3.000
58	Lk	7.89E-14	-3.869
59	Cul2	1.07E-13	-2.179
60	CG12164	1.09E-13	-4.740
61	Mfap1	1.20E-13	-1.669
62	CG13063	1.74E-13	-4.915
63	CG15731	2.04E-13	-7.146
64	CG13403	2.35E-13	-2.691
65	dyl	2.99E-13	-6.455
66	CG2962	3.79E-13	-6.592
67	CG1494	3.98E-13	-2.425
68	Trh	3.98E-13	-2.079
69	CG3690	5.52E-13	-1.850
70	Nop17l	6.20E-13	-1.076
71	CG13699	6.28E-13	-5.671
72	ple	6.53E-13	-3.207
73	CG7300	9.45E-13	-3.176
74	Ugt317A1	1.53E-12	-2.552
75	Ccp84Ac	1.78E-12	-6.208
76	e	1.78E-12	-2.260
77	CG31810	2.22E-12	-2.422
78	Lcp3	2.86E-12	-7.147
79	sage	3.13E-12	-2.121
80	CG12519	3.59E-12	-5.190
81	slbo	3.90E-12	-4.542
82	CG14147	4.62E-12	-6.470
83	CG43386	4.92E-12	-2.543
84	Ilp8	5.47E-12	-5.078
85	CG31157	6.89E-12	-3.404
86	CG31775	7.50E-12	-2.018
87	Cpr64Ab	8.12E-12	-6.970
88	CG15022	1.00E-11	-4.479
89	CG42586	1.17E-11	-1.994
90	CG7432	1.20E-11	-4.515
91	Spn100A	2.05E-11	-2.857
92	CG12964	2.33E-11	-3.154
93	CG34115	3.44E-11	-4.948
94	CG14257	4.50E-11	-4.326
95	CG43181	4.73E-11	-5.208
96	CG11147	5.46E-11	-2.366
97	CG33494	5.69E-11	-2.064
98	cysu	6.38E-11	-2.988
99	dsb	7.22E-11	-2.965
100	CG16772	7.83E-11	-4.439

Supplementary Table 2.14: Top 100 downregulated genes with the lowest padj values (T23)

Rank	Gene Symbol	padj value	log2 fc
1	mthl8	1.31E-71	-6.680
2	CG3819	4.20E-36	-5.567
3	lncRNA:CR43459	5.25E-35	-2.928
4	CG10407	1.44E-34	-3.113
5	Hsc70-2	2.69E-34	-3.593
6	Keap1	4.03E-32	-1.771
7	Grik	8.11E-31	-5.199
8	firl	2.83E-30	-3.748
9	ZnT77C	1.35E-29	-2.550
10	CG14406	4.89E-27	-8.159
11	CG1494	2.56E-26	-3.638
12	Obp99d	6.07E-25	-3.966
13	GstE1	9.67E-23	-3.079
14	CG6296	1.50E-21	-2.984
15	Cpr51A	1.54E-21	-3.989
16	Phk-3	2.25E-21	-2.926
17	Cpr65Ea	5.55E-21	-5.436
18	asRNA:CR45140	2.17E-20	-3.898
19	CG30091	1.99E-19	-4.036
20	CG33337	1.99E-19	-3.979
21	Ect3	1.29E-18	-2.490
22	CG31869	1.36E-18	-2.690
23	ThrRS	1.37E-18	-1.046
24	CG7724	1.40E-18	-2.352
25	CG9360	5.03E-18	-4.303
26	asRNA:CR45835	1.07E-17	-3.737
27	CG13159	1.30E-17	-5.306
28	Cyp6a23	1.30E-17	-2.424
29	Cul2	1.76E-17	-2.471
30	CG42807	5.11E-17	-1.941
31	CG30026	8.86E-17	-2.325
32	e	1.27E-16	-2.627
33	Spn100A	1.39E-16	-3.582
34	lncRNA:CR42646	1.76E-16	-4.892
35	asRNA:CR44179	2.77E-16	-2.376
36	CG43204	5.00E-16	-3.667
37	CG12164	6.25E-16	-5.214
38	CG31810	6.25E-16	-2.786
39	cysu	7.91E-16	-3.746
40	ckd	1.34E-15	-2.190
41	Sb	1.77E-15	-4.290
42	w	1.93E-15	-2.735
43	CG15022	3.53E-15	-5.273
44	Spn43Aa	3.90E-15	-5.054
45	scaf	3.90E-15	-2.431
46	CG7227	4.62E-15	-2.900
47	CG43386	1.88E-14	-2.832
48	asRNA:CR45822	2.97E-14	-7.730
49	Trh	5.01E-14	-2.143
50	CG4783	5.89E-14	-6.475
51	CG14186	5.89E-14	-2.586
52	CG17672	6.37E-14	-4.208
53	Cyp4p1	8.79E-14	-3.065
54	CG12769	1.42E-13	-2.550
55	Dyrk2	1.45E-13	-1.241
56	Vajk4	2.13E-13	-6.242
57	DIP-lambda	2.19E-13	-1.811
58	Ugt317A1	2.25E-13	-2.652
59	slbo	2.25E-13	-4.825
60	CG2150	2.25E-13	-4.554
61	CG9733	2.25E-13	-4.027
62	CG33346	2.50E-13	-4.160
63	CG33468	2.54E-13	-5.060
64	CG5402	2.54E-13	-3.017
65	CG34038	2.67E-13	-3.197
66	PPP1R15	2.96E-13	-0.963
67	CG13699	3.71E-13	-5.877
68	CG13659	3.71E-13	-2.983
69	asRNA:CR44029	4.51E-13	-2.822
70	CG4702	7.29E-13	-4.080
71	CG13813	7.38E-13	-5.568
72	Cad88C	7.39E-13	-2.806
73	CG33993	8.42E-13	-4.254
74	CG11147	8.52E-13	-2.565
75	Eip74EF	1.14E-12	-2.907
76	Ccp84Ac	1.43E-12	-6.313
77	Ugt37C2	1.55E-12	-2.025
78	CG15731	2.16E-12	-6.830
79	CG8027	2.40E-12	-2.017
80	CG14147	2.45E-12	-6.917
81	Vajk1	2.48E-12	-5.076
82	CG7300	2.62E-12	-3.114
83	Cpr35B	2.62E-12	-2.818
84	Ugt36E1	3.33E-12	-2.286
85	CG33494	3.37E-12	-2.170
86	b	3.80E-12	-4.112
87	NLaz	4.03E-12	-2.704
88	CG11319	4.85E-12	-1.514
89	CG42272	5.21E-12	-2.356
90	GatB	5.33E-12	-1.606
91	Gadd45	6.04E-12	-2.713
92	dsb	1.15E-11	-3.105
93	Obp99a	1.16E-11	-2.768
94	CG12519	1.25E-11	-5.090
95	Cpr64Ac	1.29E-11	-7.399
96	CG2962	1.59E-11	-6.132
97	CG34437	1.91E-11	-2.628
98	CG34115	2.55E-11	-5.302
99	lncRNA:CR43144	3.16E-11	-3.932
100	Gpb5	3.40E-11	-2.368

Supplementary Table 3.1: Top 100 upregulated genes with the lowest padj values (T10)

Rank	Gene Symbol	padj value	log2 fc
1	CG5687	9.66E-50	4.917
2	Drsl5	2.46E-31	10.500
3	CG31098	1.11E-24	2.392
4	CG42822	2.92E-21	4.245
5	CG31288	5.32E-17	3.671
6	CG18547	2.17E-15	1.632
7	CG10445	2.20E-15	3.352
8	TyrR	1.03E-14	2.293
9	AOX1	4.99E-14	1.831
10	GILT2	6.37E-14	1.162
11	lncRNA:CR44536	6.70E-14	9.477
12	CG2064	2.55E-13	2.279
13	asRNA:CR44960	5.21E-13	4.839
14	prom	6.75E-13	6.116
15	CR43105	8.83E-13	6.337
16	CG13033	1.12E-12	3.921
17	CG30059	1.13E-12	3.166
18	CG12493	3.38E-12	3.152
19	CG18278	4.13E-12	3.007
20	Est-6	8.31E-12	3.376
21	CG15515	3.98E-11	4.901
22	narya	4.45E-11	6.225
23	Ptp52F	1.08E-10	2.283
24	CG1358	2.86E-10	2.062
25	Cyp4e3	4.24E-10	3.486
26	List	5.36E-10	2.495
27	CG10353	8.25E-10	1.482
28	CG7912	8.32E-10	2.305
29	sit	1.03E-09	4.314
30	CG9449	3.66E-09	2.472
31	TM4SF	4.69E-09	2.232
32	mthl3	5.10E-09	2.550
33	CG14302	5.38E-09	4.001
34	CG31606	9.26E-09	7.636
35	Cyp12d1-d	9.41E-09	2.588
36	CG5039	1.19E-08	1.911
37	CG32237	1.67E-08	5.165
38	agt	1.74E-08	2.217
39	CR43697	1.91E-08	2.148
40	ms(3)76Ba	2.21E-08	7.790
41	Bace	3.24E-08	3.283
42	CG10953	3.47E-08	4.627
43	CR43186	4.07E-08	4.124
44	CG3270	4.25E-08	2.431
45	asRNA:CR45171	5.30E-08	3.316
46	CanA1	5.32E-08	4.174
47	TwdlP	7.00E-08	4.517
48	CG14275	7.12E-08	1.948
49	lncRNA:CR45039	7.89E-08	3.081
50	CG10031	9.47E-08	3.117
51	CG3520	1.22E-07	1.598
52	Tsp42Eb	1.38E-07	2.173
53	RpS19b	1.49E-07	1.980
54	CR42530	1.75E-07	2.566
55	Cyp28d1	1.80E-07	2.280
56	CG7714	2.17E-07	3.091
57	CG3165	3.02E-07	1.470
58	CG15369	4.23E-07	3.702
59	Mocs1	4.62E-07	1.638
60	CG14400	4.67E-07	5.894
61	TwdlH	4.71E-07	3.212
62	CG34267	5.20E-07	4.699
63	CG3635	5.47E-07	1.956
64	CG12766	7.18E-07	2.591
65	CG14196	7.20E-07	2.039
66	Odc1	7.38E-07	2.992
67	CG9521	9.59E-07	4.689
68	CG15023	9.80E-07	4.266
69	CG34166	1.04E-06	2.223
70	SP1029	1.05E-06	0.839
71	Idgf1	1.05E-06	2.818
72	sni	1.06E-06	0.916
73	CG31706	1.07E-06	3.856
74	CG6933	1.09E-06	2.721
75	Ku80	1.37E-06	1.801
76	Cyp9b2	1.57E-06	1.726
77	CG13323	1.63E-06	3.068
78	CG32335	1.68E-06	1.695
79	CG43373	1.93E-06	1.739
80	CG4563	2.01E-06	4.246
81	Muc26B	2.08E-06	3.058
82	NijA	2.26E-06	2.240
83	CG5023	2.41E-06	2.887
84	CG4730	2.51E-06	1.888
85	CG7248	2.66E-06	3.148
86	lncRNA:CR44662	2.81E-06	2.454
87	CG7465	2.94E-06	4.335
88	CG43896	3.12E-06	3.013
89	CG33272	3.37E-06	4.808
90	CG5704	3.42E-06	1.797
91	lncRNA:CR44150	3.56E-06	2.905
92	CG42319	3.65E-06	1.593
93	l(2)03659	3.74E-06	1.496
94	DNApol-zeta	3.75E-06	1.307
95	Lcp65Ac	3.83E-06	3.967
96	CG14565	4.40E-06	4.859
97	CG14445	4.74E-06	3.142
98	CG4982	4.75E-06	3.501
99	CG34005	5.65E-06	4.200
100	CG5084	6.13E-06	3.302

Supplementary Table 3.2: Top 100 upregulated genes with the lowest padj values (T11)

Rank	Gene Symbol	padj value	log2 fc
1	CG5687	7.85E-68	5.880
2	CG18547	2.19E-41	2.660
3	CG31098	3.50E-31	2.715
4	CG42822	3.14E-30	4.829
5	CG31288	7.26E-28	5.053
6	Drsl5	8.81E-28	9.519
7	lncRNA:CR44536	1.29E-27	8.828
8	AOX1	2.30E-25	2.477
9	CG2064	2.31E-23	3.021
10	CG1358	8.40E-23	3.211
11	Mocs1	5.78E-22	2.919
12	agt	1.66E-21	3.768
13	CG3448	3.27E-18	2.509
14	sni	1.43E-17	1.552
15	frm	1.81E-17	2.653
16	CG42319	2.34E-17	2.770
17	rad50	2.77E-17	2.656
18	CG10445	3.47E-16	3.415
19	CG12766	8.83E-16	5.035
20	CG7054	2.01E-15	2.356
21	CR43105	2.54E-15	7.038
22	TwdlG	3.03E-15	3.168
23	CG32694	3.67E-15	2.826
24	CG12224	5.71E-15	2.895
25	CG14907	9.34E-15	2.034
26	bnk	3.74E-14	2.096
27	Ku80	3.88E-14	2.689
28	DNApol-zeta	4.19E-14	2.031
29	LManV	5.18E-14	3.654
30	Rev7	7.28E-14	2.430
31	ms(3)76Ba	7.35E-14	7.292
32	SP1029	1.71E-13	1.205
33	TyrR	1.77E-13	2.226
34	Cyp12d1-d	3.89E-13	3.904
35	asRNA:CR45171	3.89E-13	4.756
36	CG17904	4.76E-13	1.113
37	zye	4.76E-13	2.981
38	Mco1	7.51E-13	2.717
39	Cpr62Bb	7.70E-13	3.001
40	drd	9.18E-13	2.507
41	CG31633	1.23E-12	1.339
42	Irbp	1.78E-12	2.497
43	CG2909	2.01E-12	2.586
44	asRNA:CR44960	3.79E-12	4.823
45	CG3520	4.25E-12	2.075
46	Cyp12d1-p	8.43E-12	3.879
47	CG5704	1.22E-11	2.528
48	CG30427	1.23E-11	2.196
49	CG14695	1.47E-11	3.443
50	CG4847	2.46E-11	2.434
51	Cyp9b2	2.48E-11	2.473
52	CG3165	2.56E-11	1.850
53	obe	2.82E-11	1.388
54	narya	2.82E-11	6.692
55	mthl3	3.78E-11	2.854
56	CG33281	5.83E-11	2.717
57	CG15098	9.43E-11	2.489
58	slif	1.10E-10	2.088
59	Cyp6a13	1.42E-10	1.755
60	CG10638	1.79E-10	1.634
61	CG10657	2.30E-10	2.274
62	Picot	2.67E-10	2.223
63	Fancl	2.92E-10	2.103
64	CG14196	3.89E-10	2.481
65	Prp31	5.01E-10	1.345
66	ftz-f1	5.10E-10	2.237
67	CG18278	5.80E-10	2.712
68	CG11073	6.11E-10	2.209
69	CG15649	6.56E-10	3.383
70	CG3397	7.14E-10	2.452
71	CG30059	7.15E-10	2.726
72	CG12493	7.52E-10	2.927
73	CG3008	8.35E-10	1.169
74	CG7179	1.04E-09	1.826
75	Cpr56F	1.27E-09	2.588
76	CG7080	1.65E-09	3.572
77	TM4SF	1.70E-09	2.321
78	Ptp52F	2.19E-09	2.268
79	List	2.22E-09	2.323
80	CG31102	2.31E-09	2.078
81	Dh44-R2	3.52E-09	1.535
82	CG3635	5.18E-09	2.246
83	CG6293	5.25E-09	2.701
84	RpA-70	5.31E-09	2.051
85	Proc-R	5.50E-09	2.043
86	CG5697	8.85E-09	3.552
87	Arc2	8.90E-09	1.972
88	Tb	1.22E-08	4.419
89	Est-6	1.31E-08	2.962
90	CG5568	1.42E-08	3.802
91	sit	1.44E-08	4.032
92	CR43186	1.69E-08	5.856
93	lncRNA:CR45039	1.81E-08	3.279
94	CR42530	1.93E-08	2.744
95	CG13285	2.12E-08	1.994
96	CG10031	2.27E-08	3.291
97	Bace	2.44E-08	3.285
98	CG4267	3.36E-08	2.819
99	Mur29B	3.70E-08	2.209
100	CG31606	3.86E-08	7.300

Supplementary Table 3.3: Top 100 upregulated genes with the lowest padj values (T12)

Rank	Gene Symbol	padj value	log2 fc
1	CG5687	4.05E-41	4.426
2	CG18547	9.63E-24	2.023
3	sni	2.31E-23	1.788
4	CG10445	2.29E-22	4.210
5	CG18278	5.76E-21	4.132
6	CG42822	1.42E-20	3.666
7	Cpr56F	5.34E-20	3.928
8	CG31098	8.04E-19	2.104
9	CG2064	8.04E-19	2.740
10	Mco1	8.04E-19	3.890
11	CG30059	3.97E-17	3.716
12	Fancl	8.93E-17	2.643
13	AOX1	1.00E-16	2.024
14	Picot	2.47E-16	2.825
15	CG31288	3.91E-16	3.622
16	lncRNA:CR44536	1.15E-15	8.399
17	mthl3	1.19E-15	3.651
18	CR43105	1.32E-15	5.917
19	CG11584	4.88E-15	5.627
20	slif	5.18E-15	2.461
21	rad50	1.03E-14	2.459
22	aralar1	1.31E-14	1.781
23	Cpr67Fa1	1.47E-14	5.803
24	CG1358	2.19E-14	2.409
25	CG4730	2.28E-14	2.963
26	CG4842	2.32E-14	2.960
27	tut	3.75E-14	3.947
28	CG3165	5.22E-14	2.070
29	Tsp74F	5.96E-14	1.855
30	CG6293	3.53E-13	3.253
31	narya	4.96E-13	7.707
32	Drsl5	5.01E-13	10.137
33	zye	2.86E-12	2.936
34	CG15820	4.13E-12	6.035
35	Cyp12d1-d	4.60E-12	3.521
36	CG4267	4.69E-12	3.383
37	DNApol-zeta	9.04E-12	1.852
38	CG13046	1.23E-11	3.383
39	CG31626	1.23E-11	4.622
40	CG12766	1.69E-11	3.716
41	Ku80	1.75E-11	2.436
42	asRNA:CR45171	1.96E-11	3.960
43	Calx	2.64E-11	1.485
44	Gk2	2.64E-11	2.020
45	ms(3)76Ba	2.93E-11	8.744
46	Spn88Ea	3.56E-11	1.449
47	prom	3.79E-11	4.396
48	CG14882	3.88E-11	1.478
49	CG15649	4.05E-11	4.628
50	CG17292	6.97E-11	1.337
51	Tb	8.51E-11	5.841
52	Irbp	8.68E-11	2.309
53	CG32694	1.15E-10	2.371
54	TwdlB	1.37E-10	5.091
55	Irk1	1.68E-10	1.222
56	CG6180	2.74E-10	1.841
57	CG3397	3.01E-10	2.507
58	CG2841	3.31E-10	1.372
59	Cpr62Bb	4.15E-10	2.678
60	Cyp12d1-p	4.39E-10	3.501
61	Top3alpha	4.42E-10	1.987
62	CG5704	4.61E-10	2.319
63	Cht6	5.00E-10	1.898
64	CG8563	5.07E-10	3.481
65	EbpIII	5.73E-10	2.121
66	SCAP	5.99E-10	1.626
67	EAChm	7.47E-10	4.231
68	CG34461	7.52E-10	3.764
69	SLO2	7.73E-10	1.803
70	Vajk1	8.12E-10	4.434
71	TwdlL	8.13E-10	4.604
72	LManV	9.31E-10	3.534
73	Axs	9.80E-10	1.884
74	Ipk2	1.02E-09	1.883
75	CG7080	1.09E-09	3.602
76	Cpr62Bc	1.12E-09	4.472
77	Prip	1.43E-09	2.541
78	Jafrac1	1.82E-09	2.845
79	RpA-70	2.37E-09	2.089
80	CG30427	2.44E-09	1.965
81	CG10638	2.77E-09	1.536
82	CG3520	2.77E-09	1.750
83	olf186-M	2.87E-09	2.267
84	CG15212	2.93E-09	5.710
85	TwdlW	3.21E-09	2.745
86	CG9896	5.33E-09	1.919
87	Rev7	6.09E-09	1.901
88	CG17904	6.13E-09	0.922
89	CG15279	6.25E-09	2.171
90	CG15213	6.36E-09	4.481
91	CG3552	6.63E-09	1.440
92	asRNA:CR44960	8.13E-09	4.851
93	TwdlG	8.69E-09	2.370
94	agt	9.12E-09	2.217
95	CG7601	9.67E-09	1.311
96	Ssrp	1.15E-08	1.552
97	Dip-B	1.22E-08	1.074
98	bgm	1.23E-08	1.245
99	CR43186	1.28E-08	5.068
100	Cpr67Fa2	1.60E-08	5.020

Supplementary Table 3.4: Top 100 upregulated genes with the lowest padj values (T13)

Rank	Gene Symbol	padj value	log2 fc
1	CG5687	1.15E-50	4.729
2	CG18547	1.86E-44	2.744
3	CG42822	3.46E-39	5.615
4	CG2064	3.39E-29	3.390
5	lncRNA:CR44536	1.06E-24	9.243
6	Drsl5	2.64E-23	10.139
7	CG2233	3.10E-21	3.974
8	CG31288	3.30E-17	3.175
9	CG12766	6.00E-17	4.156
10	CR43105	1.07E-16	5.807
11	ppk	1.11E-16	3.968
12	CG3165	3.35E-16	2.241
13	CG31098	5.39E-15	1.838
14	mthl3	1.17E-14	3.270
15	Cyp6a13	1.19E-14	1.792
16	asRNA:CR44030	3.89E-14	3.241
17	CG8369	1.57E-13	3.714
18	Odc1	2.03E-13	3.518
19	AOX1	7.32E-13	1.756
20	CG3397	1.28E-12	2.707
21	ms(3)76Ba	2.07E-12	7.908
22	asRNA:CR45171	2.22E-12	4.212
23	CG31706	3.08E-12	5.484
24	SCAP	6.72E-12	1.806
25	CG15515	6.72E-12	5.110
26	CG18278	8.53E-12	3.139
27	Prp31	9.62E-12	1.479
28	CG17904	1.83E-11	1.060
29	CG13694	2.00E-11	2.871
30	CG4730	2.38E-11	2.634
31	CR43186	2.98E-11	4.486
32	Ku80	4.39E-11	2.411
33	asRNA:CR44960	5.46E-11	4.637
34	CG9521	6.18E-11	6.192
35	sni	6.58E-11	1.200
36	CG2909	6.62E-11	2.459
37	CG10799	1.07E-10	3.061
38	CG10638	1.60E-10	1.657
39	narya	1.68E-10	6.528
40	Rev7	2.17E-10	2.061
41	ZnT35C	2.30E-10	1.675
42	Bin1	7.42E-10	1.685
43	CG30059	1.02E-09	2.865
44	CG31606	1.71E-09	4.812
45	Bace	1.83E-09	3.289
46	CG7465	2.24E-09	5.482
47	lncRNA:CR45039	2.28E-09	3.693
48	Cpr11A	3.56E-09	2.389
49	prom	4.07E-09	3.574
50	LManV	6.15E-09	2.546

Rank	Gene Symbol	padj value	log2 fc
51	CR43697	7.79E-09	2.172
52	CG8541	8.11E-09	3.409
53	TwdlH	8.94E-09	2.687
54	CG14569	8.99E-09	5.362
55	Cyp6w1	1.34E-08	2.553
56	CR43696	1.68E-08	3.332
57	CR42530	1.70E-08	2.785
58	CG4259	2.75E-08	2.579
59	CG14907	3.63E-08	1.524
60	FucTC	3.95E-08	2.204
61	Ugt302E1	4.39E-08	3.723
62	Cyp28d1	4.59E-08	2.421
63	CG7054	5.55E-08	1.663
64	CG5612	5.67E-08	2.476
65	DNApol-zeta	5.82E-08	1.541
66	CG12825	6.51E-08	1.783
67	Cpr78Cb	7.28E-08	2.532
68	CG32237	7.83E-08	5.015
69	lncRNA:CR45388	8.01E-08	3.397
70	CG13614	9.39E-08	2.865
71	CG14695	1.08E-07	2.749
72	CG31633	1.14E-07	1.051
73	agt	1.23E-07	2.104
74	Hml	1.23E-07	3.233
75	Lst	1.57E-07	2.258
76	TwdlP	1.85E-07	4.424
77	CG13618	2.07E-07	2.981
78	lncRNA:CR43496	2.21E-07	1.731
79	CG14302	2.58E-07	3.077
80	CG34334	2.85E-07	4.414
81	CG10445	3.09E-07	2.280
82	CG6654	3.43E-07	1.159
83	CG31832	3.65E-07	7.003
84	CG43799	4.32E-07	1.867
85	CG4982	4.49E-07	3.868
86	CG34267	4.75E-07	4.770
87	CG3635	6.25E-07	1.895
88	Cpr76Bb	8.20E-07	5.030
89	Spn88Ea	8.73E-07	1.127
90	CG11318	8.98E-07	2.168
91	CG33310	1.04E-06	2.656
92	SP1029	1.27E-06	0.848
93	lncRNA:CR44092	1.63E-06	2.632
94	CG33281	1.70E-06	1.879
95	CR45470	1.73E-06	3.281
96	CG12480	1.79E-06	4.099
97	rad50	1.92E-06	1.623
98	CG10353	1.99E-06	1.206
99	CG8736	3.49E-06	3.528
100	Edg78E	3.54E-06	2.767

Supplementary Table 3.5: Top 100 upregulated genes with the lowest padj values (T14)

Rank	Gene Symbol	padj value	log2 fc
1	CG42822	4.80E-34	5.095
2	CG18547	2.50E-27	2.148
3	CG2233	8.04E-25	4.013
4	ppk	8.04E-25	4.474
5	lncRNA:CR44536	7.13E-24	9.017
6	CG10445	3.36E-22	4.368
7	CG31706	1.50E-21	4.424
8	CG2064	5.51E-21	2.877
9	CG3165	5.59E-21	2.538
10	CG8541	6.63E-20	3.854
11	CG14014	2.68E-19	3.088
12	DNApol-zeta	1.10E-18	2.379
13	CG18278	1.42E-17	3.779
14	CG30059	4.36E-17	3.814
15	CG13694	8.41E-15	2.859
16	CG5687	1.73E-14	2.450
17	mthl3	1.73E-14	3.116
18	CG4730	6.19E-14	2.894
19	Odc1	7.64E-14	3.151
20	CG3397	1.15E-13	2.845
21	CR43105	3.42E-13	5.113
22	asRNA:CR45171	6.69E-13	4.188
23	Drsl5	6.69E-13	10.097
24	CG4842	1.87E-12	2.560
25	ms(3)76Ba	2.14E-12	9.278
26	CG6300	2.34E-12	4.618
27	Rev7	2.54E-12	2.282
28	agt	2.54E-12	2.728
29	Ugt302E1	3.14E-12	3.932
30	sni	5.07E-12	1.246
31	Bin1	5.29E-12	1.846
32	SCAP	7.47E-12	1.784
33	tut	7.61E-12	3.504
34	narya	1.49E-11	6.086
35	CG7054	2.53E-11	1.961
36	CG10799	3.02E-11	3.157
37	asRNA:CR44030	5.35E-11	2.940
38	CR42530	1.46E-10	3.179
39	CG6293	1.65E-10	2.898
40	Cpr11A	3.09E-10	2.485
41	CG10910	3.27E-10	4.034
42	CG33301	3.35E-10	5.327
43	CG12766	3.69E-10	2.856
44	Irbp	3.92E-10	2.248
45	CG11318	1.44E-09	2.390
46	rad50	1.61E-09	1.972
47	CG6763	1.73E-09	2.571
48	CG10560	1.86E-09	3.308
49	Cpr67B	1.87E-09	3.685
50	Lst	1.91E-09	2.415

Rank	Gene Symbol	padj value	log2 fc
51	CR43696	3.56E-09	2.695
52	CG10638	4.18E-09	1.528
53	CG32104	4.59E-09	1.344
54	CG9877	4.96E-09	5.226
55	CG7290	5.21E-09	4.941
56	CG31098	5.31E-09	1.399
57	CG11659	7.49E-09	3.525
58	Cyp12a4	7.49E-09	3.656
59	CG34334	1.21E-08	4.785
60	CG31288	1.35E-08	2.069
61	LManVI	1.37E-08	4.023
62	CG7953	1.45E-08	4.423
63	CR45470	1.47E-08	3.458
64	lncRNA:CR45039	1.58E-08	3.517
65	Ku80	1.64E-08	2.091
66	CG12825	1.98E-08	1.805
67	Cyp28d1	2.25E-08	2.399
68	CR43186	3.49E-08	3.516
69	asRNA:CR44960	3.50E-08	3.850
70	Lip3	4.79E-08	4.267
71	CG15820	4.79E-08	4.466
72	CG11601	4.81E-08	1.113
73	MED28	8.63E-08	1.267
74	Ugt37A2	1.06E-07	3.650
75	DNApol-eta	1.08E-07	1.817
76	CG12863	1.15E-07	1.553
77	CG9682	1.28E-07	4.634
78	CG14907	1.42E-07	1.444
79	Cyp6a13	1.80E-07	1.179
80	CG6283	1.90E-07	4.452
81	Hml	1.92E-07	3.084
82	Cyp12d1-d	1.94E-07	2.530
83	CG5039	2.04E-07	1.898
84	CG13658	2.25E-07	2.912
85	Rh50	2.27E-07	2.169
86	FucTC	2.52E-07	2.052
87	lncRNA:CR44662	2.59E-07	2.517
88	CG31832	2.59E-07	7.020
89	Jon99Fi	2.79E-07	5.545
90	Jon25Bi	2.86E-07	5.151
91	CG17191	5.10E-07	4.244
92	CG6910	5.29E-07	3.549
93	CG14302	5.60E-07	2.636
94	CG17134	5.66E-07	3.665
95	CG45207	5.72E-07	4.255
96	CG14882	5.91E-07	1.165
97	CG3552	5.91E-07	1.276
98	RpA-70	6.20E-07	1.795
99	Cyp6w1	6.99E-07	2.187
100	CG18180	7.12E-07	6.170

Supplementary Table 3.6: Top 100 upregulated genes with the lowest padj values (T15)

Rank	Gene Symbol	padj value	log2 fc
1	CG2233	9.47E-39	4.925
2	ppk	1.16E-34	5.430
3	CG42822	1.47E-33	4.567
4	CG18547	2.30E-27	2.129
5	CG2064	1.67E-26	3.231
6	CG11659	3.50E-23	5.851
7	CG6300	9.64E-23	6.128
8	CG31706	2.03E-22	3.610
9	CG12766	1.20E-20	3.981
10	CG10560	2.64E-20	4.066
11	CG12643	8.39E-20	2.912
12	CG8541	8.89E-20	3.389
13	CG17549	2.82E-19	3.691
14	Drsl5	6.15E-19	7.234
15	Bin1	1.02E-18	2.310
16	CG32241	1.52E-18	6.256
17	Ugt302E1	1.14E-17	6.677
18	CG7299	2.13E-17	4.084
19	Jon74E	2.23E-17	4.819
20	CG11413	2.82E-17	3.856
21	SCAP	3.12E-17	2.151
22	CG11852	2.17E-16	5.811
23	CG6910	6.89E-16	5.494
24	CG10799	1.04E-15	3.843
25	CG40198	1.15E-15	5.176
26	CG15597	2.11E-15	3.966
27	Ugt37A3	2.78E-15	4.220
28	CG11318	9.49E-15	2.774
29	CG3165	1.06E-14	2.134
30	CG17191	2.84E-14	5.388
31	asRNA:CR44030	3.28E-14	3.610
32	Odc1	4.86E-14	2.976
33	CR43105	4.86E-14	4.164
34	CR43697	5.87E-14	2.739
35	sni	5.93E-14	1.337
36	CG10445	1.13E-13	3.486
37	CG31103	1.25E-13	3.831
38	lncRNA:CR44536	1.30E-13	10.088
39	CG10513	2.22E-13	3.062
40	CG33301	2.95E-13	4.895
41	CG7054	4.13E-13	2.099
42	CG31104	8.14E-13	4.452
43	CG10638	1.09E-12	1.804
44	CG31288	1.09E-12	2.409
45	Cyp6a13	1.60E-12	1.486
46	Cpr67B	2.66E-12	4.077
47	CG17738	4.23E-12	4.400
48	CG31098	4.33E-12	1.612
49	CG31778	5.22E-12	2.321
50	yellow-d	7.29E-12	2.886

Rank	Gene Symbol	padj value	log2 fc
51	CG3397	8.07E-12	2.471
52	Mdr50	9.33E-12	2.801
53	CR45470	1.09E-11	4.766
54	asRNA:CR44960	1.11E-11	5.073
55	Cyp12a4	1.21E-11	4.138
56	Gbp3	1.59E-11	2.571
57	Cpr11B	1.73E-11	4.414
58	CG12825	1.89E-11	2.072
59	Rev7	2.68E-11	2.144
60	Cyp6w1	3.14E-11	2.712
61	lncRNA:CR45039	4.46E-11	4.151
62	CG9682	4.68E-11	5.290
63	CG5767	5.55E-11	6.039
64	CG9877	8.73E-11	5.528
65	Cht5	9.46E-11	2.806
66	DNApol-zeta	1.37E-10	1.779
67	mthl3	1.71E-10	2.501
68	CG10562	1.73E-10	2.228
69	CG5810	1.73E-10	3.048
70	bgm	1.90E-10	1.374
71	CG30047	1.96E-10	3.012
72	Cyp28d1	2.18E-10	2.648
73	Cyp4d14	2.24E-10	3.336
74	CG10912	2.38E-10	2.310
75	CG3604	2.93E-10	2.208
76	Cpr49Ad	2.98E-10	6.358
77	asRNA:CR45206	3.02E-10	5.268
78	CG14219	3.34E-10	3.794
79	Crys	3.36E-10	3.687
80	CG31321	4.15E-10	2.982
81	CG7290	4.93E-10	5.166
82	CR43696	6.25E-10	2.427
83	CR43186	6.96E-10	3.212
84	CG11585	7.97E-10	4.989
85	lncRNA:CR44662	8.42E-10	2.800
86	CG43799	8.90E-10	2.183
87	CG10559	9.36E-10	2.390
88	thetaTry	9.66E-10	3.882
89	CG6271	1.08E-09	5.027
90	CG34026	1.22E-09	3.458
91	CG42808	1.29E-09	2.527
92	l(3)mbn	1.32E-09	5.598
93	CG14275	1.74E-09	2.093
94	CG10910	1.88E-09	3.653
95	CG13694	1.97E-09	2.164
96	CG31272	2.25E-09	1.748
97	CG30059	3.06E-09	2.885
98	CG4842	3.38E-09	2.043
99	asRNA:CR45171	3.56E-09	3.613
100	CG4607	3.61E-09	2.440

Supplementary Table 3.7: Top 100 upregulated genes with the lowest padj values (T16)

Rank	Gene Symbol	padj value	log2 fc
1	ppk	1.13E-28	5.242
2	Ugt302E1	2.06E-24	4.851
3	CG30059	3.96E-22	4.812
4	CG3165	6.88E-21	2.550
5	CG18278	7.69E-21	4.226
6	CG12643	1.45E-20	2.974
7	Bin1	1.80E-20	2.424
8	CG6300	4.47E-20	5.660
9	CG11659	2.06E-19	5.154
10	CG10445	5.96E-18	3.832
11	tut	3.26E-17	4.355
12	CG42822	7.27E-17	3.285
13	Irbp	4.26E-16	2.855
14	CG2064	7.59E-16	2.497
15	lncRNA:CR44536	3.41E-15	10.068
16	DNApol-zeta	4.03E-15	2.128
17	Tmem18	1.13E-14	2.318
18	sni	1.34E-14	1.349
19	asRNA:CR44030	2.37E-14	3.475
20	Drsl5	2.63E-14	6.305
21	Ssrp	4.77E-14	1.998
22	Fen1	5.97E-14	2.349
23	MED28	7.87E-14	1.747
24	CG32104	9.55E-14	1.708
25	CG32243	1.03E-13	2.615
26	Nup43	1.57E-13	1.995
27	Mes4	2.60E-13	2.293
28	CG12863	3.61E-13	2.076
29	CG1239	4.95E-13	3.334
30	CG17977	1.13E-12	1.984
31	CG7011	1.48E-12	1.273
32	Hira	1.69E-12	1.995
33	JMJD7	2.34E-12	2.335
34	Mat1	4.06E-12	2.213
35	Rev7	4.66E-12	2.263
36	CG2233	4.83E-12	2.534
37	RnrL	7.15E-12	2.579
38	CG11601	7.85E-12	1.376
39	Cpsf73	1.17E-11	1.891
40	SCAP	1.33E-11	1.752
41	Cdc23	1.50E-11	2.487
42	mtSSB	2.18E-11	1.929
43	CG31249	2.39E-11	1.561
44	CG10638	2.57E-11	1.698
45	Ku80	2.72E-11	2.423
46	Phax	3.01E-11	1.449
47	bip1	3.03E-11	3.236
48	DNApol-delta	4.05E-11	2.935
49	CG4730	4.20E-11	2.562
50	CG15443	4.27E-11	1.824
51	Pol32	4.74E-11	2.780
52	Rad9	5.80E-11	2.405
53	RpA-70	1.29E-10	2.226
54	DNApol-alpha60	1.44E-10	2.853
55	Psf3	1.48E-10	2.188
56	beta4GalNAcTB	1.77E-10	1.505
57	Sil1	1.98E-10	1.417
58	lncRNA:CR42861	1.98E-10	2.103
59	mdlc	2.30E-10	2.967
60	CG1890	2.49E-10	1.996
61	TORIP	2.84E-10	1.761
62	CG7601	4.59E-10	1.409
63	CR45470	4.76E-10	4.156
64	agt	4.98E-10	2.453
65	asRNA:CR45171	5.91E-10	3.657
66	CG3226	6.23E-10	2.348
67	beta4GalT7	6.70E-10	2.101
68	rumi	7.61E-10	2.797
69	Nup133	8.20E-10	1.459
70	rad50	8.76E-10	1.983
71	narya	1.19E-09	7.188
72	CG3552	1.21E-09	1.508
73	EMC2A	1.21E-09	1.028
74	thoc6	1.36E-09	2.003
75	CG14881	1.91E-09	2.299
76	Vlet	2.06E-09	1.384
77	DNApol-eta	2.26E-09	1.993
78	Orc4	2.68E-09	2.070
79	CG5535	3.06E-09	1.453
80	CG9344	3.06E-09	1.833
81	DNAlig1	3.31E-09	2.655
82	slx1	3.32E-09	2.812
83	mus201	3.34E-09	1.541
84	Ugt37A3	3.52E-09	2.829
85	dare	3.85E-09	1.246
86	CG3520	4.41E-09	1.737
87	ms(3)76Ba	4.55E-09	8.148
88	CG5568	5.04E-09	3.868
89	mei-41	5.63E-09	1.991
90	APC4	5.89E-09	2.138
91	CG14882	6.91E-09	1.285
92	CG14230	7.41E-09	1.528
93	hay	8.38E-09	1.128
94	Usp5	8.93E-09	1.785
95	lncRNA:CR45039	9.46E-09	3.638
96	PIG-C	1.01E-08	2.004
97	stet	1.01E-08	3.209
98	Nmt	1.02E-08	2.324
99	CG5986	1.15E-08	1.343
100	Sem1	1.18E-08	1.745

Supplementary Table 3.8: Top 100 upregulated genes with the lowest padj values (T17)

Rank	Gene Symbol	padj value	log2 fc
1	ppk	5.47E-32	5.556
2	Ugt302E1	4.22E-28	5.547
3	lncRNA:CR44536	6.66E-25	9.201
4	CG3165	1.04E-22	2.646
5	CG12643	2.92E-20	2.962
6	CG4730	6.13E-20	3.504
7	CG10445	7.85E-17	3.799
8	CG6398	3.92E-16	1.248
9	asRNA:CR44030	5.40E-16	4.041
10	Bin1	8.20E-16	2.129
11	CG5316	1.31E-15	1.366
12	CG30059	4.29E-15	3.389
13	Drsl5	1.04E-14	6.384
14	CG18278	1.54E-14	3.271
15	Ugt37A3	1.54E-14	3.575
16	CG2233	1.77E-14	2.807
17	CG6300	7.52E-14	4.575
18	CG11659	8.02E-14	4.223
19	CG42822	1.53E-13	2.987
20	CR43105	2.18E-13	3.501
21	tut	2.43E-13	3.554
22	CG3552	5.60E-13	1.791
23	ms(3)76Ba	7.07E-13	9.471
24	narya	2.49E-12	6.259
25	CR43697	5.72E-12	2.890
26	CG2064	8.03E-12	2.146
27	lncRNA:CR45039	1.27E-11	4.894
28	CR42530	2.78E-11	3.251
29	CR45470	4.08E-11	4.389
30	CG12863	4.93E-11	1.843
31	asRNA:CR45171	4.95E-11	3.613
32	CG7011	5.42E-11	1.164
33	lncRNA:CR42861	9.35E-11	2.120
34	Cyp28d2	1.34E-10	2.201
35	Odc2	1.36E-10	3.361
36	CR44003	4.27E-10	3.783
37	lncRNA:CR43496	7.17E-10	1.969
38	CR43186	8.35E-10	2.927
39	CG17177	1.68E-09	5.234
40	asRNA:CR44960	3.47E-09	4.933
41	slx1	4.17E-09	2.718
42	CG7601	7.51E-09	1.308
43	rtet	9.28E-09	1.591
44	CG5986	1.19E-08	1.345
45	sni	1.20E-08	1.015
46	l(3)mbn	1.31E-08	2.858
47	SCAP	4.62E-08	1.449
48	CG10560	4.62E-08	1.805
49	Drsl2	5.55E-08	6.253
50	Arp6	5.56E-08	0.709
51	Cpr49Ad	5.70E-08	3.861
52	CG31104	5.90E-08	2.468
53	CG9344	7.22E-08	1.689
54	Tmem18	8.01E-08	1.615
55	CG17977	8.87E-08	1.495
56	CG7694	1.02E-07	1.669
57	CG18814	1.03E-07	2.311
58	Phax	1.29E-07	1.167
59	CG15432	1.31E-07	0.979
60	CG31633	1.35E-07	1.020
61	CG2921	2.56E-07	0.876
62	CG5323	2.92E-07	1.234
63	CG10638	3.38E-07	1.335
64	SmE	3.62E-07	1.619
65	CG11601	3.87E-07	1.031
66	RhoGEF4	4.06E-07	2.461
67	CG32243	4.14E-07	1.825
68	CG15443	4.56E-07	1.419
69	Qtzl	4.74E-07	1.147
70	Vlet	4.95E-07	1.171
71	CG30496	4.95E-07	1.338
72	CG3631	4.95E-07	1.208
73	mal	5.28E-07	1.223
74	c(3)G	5.49E-07	1.997
75	l(1)10Bb	7.52E-07	1.335
76	shv	7.96E-07	0.605
77	MED11	9.55E-07	1.216
78	Cyp4e3	9.95E-07	3.213
79	SCCRO4	1.27E-06	1.583
80	lncRNA:CR45225	1.41E-06	4.222
81	Sil1	1.87E-06	1.077
82	CG5039	2.23E-06	1.701
83	CG7483	2.34E-06	0.982
84	Usp39	2.66E-06	1.494
85	Uhg3	2.67E-06	1.225
86	EMC2A	2.85E-06	0.809
87	dare	3.04E-06	1.013
88	TORIP	3.47E-06	1.323
89	CG6066	3.53E-06	1.684
90	adp	3.68E-06	0.849
91	CG7137	4.09E-06	0.956
92	CG5882	4.32E-06	2.048
93	CG6910	4.57E-06	3.085
94	CG17904	4.59E-06	0.750
95	spn-F	4.70E-06	1.280
96	CG9586	4.71E-06	0.952
97	CG11777	4.88E-06	1.374
98	mRpS18B	4.91E-06	0.865
99	mia	5.63E-06	1.214
100	CG5704	5.63E-06	1.716

Supplementary Table 3.9: Top 100 upregulated genes with the lowest padj values (T18)

Rank	Gene Symbol	padj value	log2 fc
1	Ugt302E1	5.07E-33	5.623
2	ppk	3.10E-28	4.726
3	lncRNA:CR44536	8.80E-27	8.700
4	CG3165	7.64E-22	2.613
5	Bin1	8.05E-22	2.519
6	Ssrp	4.89E-20	2.392
7	CG18278	4.96E-19	4.082
8	CG30059	1.50E-18	4.064
9	sni	2.41E-18	1.521
10	Irbp	7.88E-18	3.020
11	MED28	7.88E-18	2.025
12	CG10445	1.31E-17	3.956
13	DNApol-zeta	1.64E-17	2.296
14	beta4GalT7	2.54E-17	2.923
15	RpA-70	3.94E-17	2.848
16	Mes4	5.23E-17	2.677
17	CG12643	8.07E-17	2.664
18	SmD3	8.99E-17	2.503
19	Usp5	1.24E-16	2.505
20	CG9344	1.94E-16	2.587
21	CG31249	1.94E-16	1.946
22	CG32104	3.03E-16	1.903
23	Orc4	1.42E-15	2.747
24	MED16	1.82E-15	2.130
25	Cpsf73	2.10E-15	2.197
26	Psf3	3.47E-15	2.682
27	Ugt37A3	5.71E-15	3.563
28	JMJD7	1.60E-14	2.548
29	tut	3.07E-14	3.793
30	beta4GalNAcTB	3.35E-14	1.762
31	asRNA:CR44030	5.60E-14	3.375
32	Cdc23	5.60E-14	2.739
33	Alg9	5.60E-14	2.677
34	Mat1	6.39E-14	2.375
35	RPA2	8.27E-14	3.595
36	Hira	8.29E-14	2.104
37	DNApol-alpha60	8.94E-14	3.297
38	Ndc80	1.40E-13	3.805
39	LSm7	1.99E-13	1.816
40	Ku80	2.90E-13	2.655
41	CG7544	3.15E-13	2.895
42	DNAlig1	3.57E-13	3.214
43	mtSSB	3.57E-13	2.078
44	PCNA	3.82E-13	3.456
45	Phax	6.23E-13	1.552
46	CG11601	8.38E-13	1.439
47	CG31075	1.16E-12	2.309
48	SmE	1.48E-12	2.170
49	CG4730	1.93E-12	2.730
50	Arp6	2.95E-12	0.895
51	Fen1	2.97E-12	2.170
52	CG3397	3.03E-12	2.648
53	thoc6	3.03E-12	2.278
54	dgt3	4.18E-12	4.289
55	Nup43	4.18E-12	1.863
56	Pol31	7.14E-12	3.544
57	CG6654	8.34E-12	1.476
58	Xpd	8.86E-12	2.677
59	grsm	8.88E-12	2.972
60	slx1	9.68E-12	3.311
61	CG31251	1.06E-11	2.756
62	Mis12	1.17E-11	3.294
63	CG12863	1.25E-11	1.921
64	CG2064	1.63E-11	2.089
65	Nup133	1.63E-11	1.578
66	asRNA:CR45171	1.91E-11	3.984
67	CG17266	2.51E-11	2.714
68	Cdk2	3.05E-11	3.517
69	RnrS	3.05E-11	2.987
70	Ssb-c31a	3.05E-11	1.957
71	tam	3.17E-11	2.081
72	Mlh1	3.27E-11	2.188
73	ms(3)76Ba	3.30E-11	8.564
74	dre4	3.51E-11	1.927
75	Drsl5	3.60E-11	5.097
76	Caf1-105	3.65E-11	4.169
77	Dsor1	4.78E-11	2.058
78	lncRNA:CR45039	6.31E-11	4.609
79	CG17470	6.44E-11	4.604
80	RnrL	6.81E-11	2.436
81	Npl4	1.06E-10	1.718
82	Orc2	1.10E-10	3.879
83	Prp31	1.11E-10	1.388
84	CR31953	1.14E-10	4.267
85	DNApol-delta	1.40E-10	2.828
86	CG17977	1.46E-10	1.860
87	dpa	1.49E-10	2.486
88	SmD1	1.98E-10	2.822
89	tos	2.38E-10	4.760
90	CG8420	2.42E-10	3.216
91	Exd2	2.45E-10	1.291
92	CG12942	2.57E-10	2.622
93	CG10336	2.61E-10	3.971
94	Sem1	2.64E-10	1.897
95	DNApol-alpha73	2.71E-10	3.583
96	janA	2.93E-10	1.888
97	SmF	3.25E-10	2.289
98	CG12321	3.31E-10	1.524
99	ida	3.45E-10	3.000
100	Pop2	3.71E-10	1.473

Supplementary Table 3.10: Top 100 upregulated genes with the lowest padj values (T19)

Rank	Gene Symbol	padj value	log2 fc
1	ppk	2.38E-26	5.295
2	Ugt302E1	3.67E-26	7.029
3	CG12643	1.77E-23	3.204
4	Bin1	3.12E-23	2.614
5	lncRNA:CR44536	4.57E-23	8.965
6	CG2064	1.10E-19	2.803
7	CG10638	2.13E-17	2.154
8	bnk	2.45E-16	2.859
9	CG10445	2.46E-16	4.214
10	CG3165	5.02E-16	2.244
11	CG3397	1.07E-15	3.363
12	Ku80	3.20E-15	2.969
13	GstD5	3.71E-15	5.036
14	CG14014	3.67E-14	2.166
15	asRNA:CR44030	9.44E-14	4.200
16	CG15784	1.08E-13	3.543
17	RpA-70	5.03E-13	2.496
18	CG4730	8.23E-13	2.848
19	Ugt37A3	1.49E-12	3.232
20	Alg9	1.78E-12	2.649
21	CG18278	2.06E-12	3.225
22	CG9344	2.37E-12	2.394
23	CG11659	3.20E-12	3.923
24	tut	5.12E-12	3.709
25	CG6770	9.27E-12	1.448
26	CG3552	9.71E-12	1.782
27	CG6300	9.71E-12	4.169
28	Cyp6a17	1.23E-11	2.115
29	Tmem18	1.43E-11	2.204
30	CG30059	1.69E-11	3.128
31	MED28	2.45E-11	1.674
32	rad50	2.69E-11	2.175
33	CG7601	3.15E-11	1.516
34	CG6654	4.73E-11	1.460
35	DNApol-zeta	4.78E-11	1.819
36	CG17977	7.08E-11	2.061
37	Ssb-c31a	7.43E-11	2.008
38	CG11601	1.18E-10	1.320
39	Ssrp	1.19E-10	1.741
40	ms(3)76Ba	1.19E-10	8.503
41	CG17177	1.76E-10	5.886
42	CR43697	2.02E-10	2.954
43	Bsg25A	2.12E-10	1.722
44	CG3085	2.21E-10	2.921
45	Irbp	4.12E-10	2.262
46	narya	4.35E-10	5.958
47	CG32104	6.02E-10	1.494
48	Mes4	6.24E-10	2.050
49	lncRNA:CR45039	6.24E-10	4.626
50	CG13694	6.25E-10	2.399
51	CG32243	7.71E-10	2.215
52	Prat2	8.34E-10	2.083
53	CG5157	1.42E-09	1.727
54	Grip71	1.44E-09	4.372
55	Ndc80	1.72E-09	3.227
56	Surf1	1.86E-09	1.889
57	CG34166	1.86E-09	2.459
58	beta4GalNAcTB	2.04E-09	1.444
59	Vlet	2.31E-09	1.420
60	MED16	2.92E-09	1.649
61	beta4GalT7	3.08E-09	2.137
62	Pif1A	3.41E-09	2.635
63	tam	4.02E-09	1.909
64	CR45470	4.17E-09	4.173
65	CG5039	4.36E-09	2.317
66	Drsl2	4.36E-09	6.804
67	trsn	4.56E-09	1.857
68	asRNA:CR45171	5.22E-09	3.494
69	Ipk1	6.20E-09	1.595
70	Rev7	6.69E-09	2.052
71	TORIP	6.94E-09	1.651
72	thoc6	7.83E-09	1.958
73	CG14881	8.24E-09	2.341
74	LSm7	8.91E-09	1.496
75	Psf3	9.00E-09	2.023
76	Cdc23	9.35E-09	2.167
77	CG12224	1.26E-08	2.105
78	CR44003	1.69E-08	3.732
79	Hira	1.84E-08	1.653
80	CG12863	1.86E-08	1.655
81	CG5568	1.87E-08	3.757
82	Usp5	2.16E-08	1.768
83	slx1	3.04E-08	2.779
84	asRNA:CR44960	3.22E-08	5.005
85	CG5882	3.24E-08	3.103
86	CG15547	3.24E-08	2.314
87	Pop2	4.06E-08	1.329
88	agt	4.24E-08	2.212
89	RPA2	4.80E-08	2.718
90	CG31075	5.68E-08	1.843
91	CG34348	5.74E-08	1.426
92	Pol31	6.18E-08	2.900
93	CG7011	6.26E-08	1.003
94	CG17266	6.45E-08	2.320
95	CG15443	6.46E-08	1.531
96	CR42530	6.48E-08	2.926
97	CG33213	6.89E-08	2.860
98	krimp	6.99E-08	4.196
99	Prp31	7.53E-08	1.194
100	mdlc	8.63E-08	2.544

Supplementary Table 3.11: Top 100 upregulated genes with the lowest padj values (T20)

Rank	Gene Symbol	padj value	log2 fc
1	Ugt302E1	5.72E-41	7.413
2	lncRNA:CR44536	1.23E-38	8.758
3	ppk	8.94E-33	5.586
4	CG12643	1.88E-31	3.675
5	CG2064	1.63E-27	3.308
6	CG34166	1.57E-20	3.678
7	Cyp6a17	5.09E-19	2.720
8	Odc2	2.21E-17	3.837
9	Cyp28d2	4.11E-16	2.694
10	Ugt37A1	7.47E-16	4.062
11	GstD5	1.48E-15	5.108
12	mrt	2.29E-15	1.598
13	asRNA:CR44030	2.10E-14	4.628
14	CG6770	2.81E-14	1.611
15	CR43697	5.22E-14	3.489
16	CG31300	8.65E-14	2.842
17	CG5157	8.77E-14	2.072
18	lncRNA:CR44662	1.47E-13	3.419
19	CG3397	2.28E-13	2.673
20	Reg-2	2.63E-13	1.966
21	CG13694	3.59E-13	2.812
22	CG8654	4.11E-13	2.161
23	CG17977	6.07E-13	2.272
24	Cat	6.25E-13	1.134
25	Phae2	6.25E-13	2.575
26	Bin1	9.22E-13	1.925
27	CG13078	1.74E-12	4.605
28	CG3552	2.99E-12	1.830
29	CG8665	3.50E-12	2.760
30	CG34105	3.72E-12	8.519
31	CG4335	7.47E-12	3.166
32	CG10445	8.88E-12	3.212
33	CG42319	9.08E-12	2.159
34	CG6654	2.05E-11	1.479
35	Ugt37A3	3.91E-11	3.015
36	Cyp4e3	4.10E-11	4.746
37	DNApol-zeta	6.68E-11	1.799
38	CG3165	7.00E-11	1.844
39	CG30414	7.08E-11	2.257
40	CG31683	9.29E-11	1.207
41	Drsl2	1.00E-10	6.622
42	CG15784	1.02E-10	3.123
43	lncRNA:CR45039	1.29E-10	5.578
44	lncRNA:CR43496	1.62E-10	2.260
45	Prat2	1.91E-10	2.114
46	CG31104	3.31E-10	2.488
47	Iyd	3.96E-10	2.660
48	CG31075	4.13E-10	2.067
49	ms(3)76Ba	4.13E-10	8.256
50	CG34215	4.43E-10	2.190

Rank	Gene Symbol	padj value	log2 fc
51	Thor	8.47E-10	1.493
52	CG12224	1.03E-09	2.213
53	Pif1A	1.06E-09	2.597
54	CG34220	1.08E-09	2.326
55	Lip3	1.39E-09	4.379
56	CG12491	1.39E-09	9.465
57	CG18858	1.42E-09	1.041
58	twz	1.65E-09	1.841
59	CG10638	1.91E-09	1.566
60	CG17177	2.00E-09	8.124
61	CG10936	2.24E-09	1.654
62	CG31633	2.32E-09	1.185
63	lncRNA:CR43493	3.63E-09	2.498
64	CG5882	3.66E-09	3.741
65	FucTC	4.92E-09	2.279
66	GstD6	6.65E-09	2.341
67	RabX1	9.37E-09	1.223
68	CG30059	1.05E-08	2.697
69	CG7912	1.06E-08	2.046
70	CG33189	2.03E-08	2.512
71	Cyp12d1-d	2.04E-08	2.327
72	CG16704	3.29E-08	2.736
73	CG11601	3.41E-08	1.139
74	asRNA:CR44960	4.73E-08	4.735
75	CG32335	5.04E-08	1.969
76	CG15547	5.43E-08	2.251
77	CG18278	5.69E-08	2.502
78	Lectin-galC1	5.78E-08	2.986
79	CG5316	6.31E-08	0.979
80	asRNA:CR45136	1.01E-07	2.841
81	Lerp	1.18E-07	1.071
82	CG5009	1.50E-07	1.013
83	Ku80	1.50E-07	1.970
84	CG18563	1.63E-07	2.912
85	Npc2c	1.75E-07	3.396
86	lectin-24A	1.77E-07	3.114
87	Sardh	2.05E-07	1.898
88	asRNA:CR45171	2.75E-07	3.136
89	CG31817	3.07E-07	2.815
90	Pepck2	4.14E-07	2.122
91	Ugt36D1	4.20E-07	1.433
92	Mics1	4.20E-07	3.228
93	CG15369	4.54E-07	3.005
94	CG12522	4.54E-07	5.449
95	CR43105	4.66E-07	2.197
96	lncRNA:CR45517	4.83E-07	1.702
97	LManIII	5.04E-07	3.623
98	narya	5.56E-07	6.111
99	sni	5.62E-07	0.921
100	mthl13	6.11E-07	2.927

Supplementary Table 3.12: Top 100 upregulated genes with the lowest padj values (T21)

Rank	Gene Symbol	padj value	log2 fc
1	Ugt302E1	1.19E-33	6.930
2	ppk	1.58E-30	6.079
3	CG12643	8.59E-27	3.403
4	CG2064	2.56E-26	3.234
5	lncRNA:CR44536	4.28E-26	9.532
6	Cyp6a17	1.88E-24	3.090
7	Odc2	4.48E-22	4.641
8	Cyp28d2	1.47E-20	3.052
9	CG8665	1.93E-19	3.521
10	CG6398	8.38E-19	1.354
11	ms(3)76Ba	2.00E-18	7.810
12	CG3165	4.60E-17	2.307
13	CG5316	1.29E-13	1.294
14	asRNA:CR44030	1.35E-13	4.472
15	Bin1	8.79E-13	1.933
16	CG13962	1.04E-12	4.178
17	CG10445	1.23E-12	3.339
18	CG4730	1.34E-12	2.825
19	GstD5	1.81E-12	4.571
20	Ugt37A3	1.97E-12	3.224
21	lncRNA:CR45039	3.39E-12	4.659
22	CG42319	8.06E-12	2.182
23	CG31300	8.06E-12	2.608
24	lncRNA:CR44662	8.06E-12	3.155
25	asRNA:CR45171	1.25E-11	4.244
26	CG31683	1.34E-11	1.257
27	CG18858	1.79E-11	1.145
28	upSET	1.84E-11	1.504
29	Iyd	2.18E-11	2.843
30	CG8654	2.82E-11	2.012
31	CG4335	3.23E-11	3.203
32	CR43105	6.04E-11	2.957
33	CG34166	8.82E-11	2.657
34	CG16704	1.08E-10	3.248
35	CG30059	1.27E-10	3.030
36	CR43697	1.39E-10	2.799
37	Cyp4e3	1.94E-10	3.764
38	Ugt36D1	1.95E-10	1.767
39	CG34215	3.82E-10	2.217
40	lectin-24A	3.88E-10	3.816
41	CG17977	4.99E-10	1.942
42	Cyp12d1-d	1.45E-09	2.555
43	CG6665	1.72E-09	1.214
44	mrt	1.76E-09	1.252
45	CG10638	1.78E-09	1.578
46	CG6770	1.84E-09	1.307
47	CG15432	2.18E-09	1.123
48	CG5882	2.37E-09	4.176
49	l(2)gl	2.80E-09	1.192
50	CR45470	4.00E-09	4.010
51	sni	4.27E-09	1.073
52	CG31633	4.27E-09	1.169
53	mv	4.71E-09	1.552
54	CG43110	5.26E-09	2.412
55	Phae2	5.32E-09	2.136
56	lncRNA:CR43496	5.52E-09	2.017
57	CG12491	5.53E-09	8.329
58	CG3397	5.58E-09	2.173
59	CG15547	5.84E-09	2.433
60	CG34105	6.28E-09	9.376
61	CG2493	6.98E-09	1.839
62	Cyp6d2	7.39E-09	2.515
63	Sardh	7.41E-09	2.106
64	Tmem131	8.26E-09	1.751
65	CG30414	1.34E-08	1.981
66	narya	1.36E-08	5.337
67	CG6201	1.64E-08	1.569
68	CG31104	1.75E-08	2.262
69	Dab	2.02E-08	0.855
70	CG11475	2.14E-08	3.549
71	CG8155	2.44E-08	1.668
72	Dlip3	2.75E-08	1.165
73	cno	2.76E-08	1.423
74	CG18278	3.12E-08	2.529
75	CG10936	3.88E-08	1.544
76	CG15784	4.85E-08	2.698
77	RhoGEF4	4.93E-08	2.656
78	FucTC	5.39E-08	2.153
79	rtet	5.61E-08	1.526
80	asRNA:CR44960	6.49E-08	3.830
81	Qtzl	7.19E-08	1.238
82	poe	7.48E-08	1.283
83	unk	7.79E-08	1.119
84	qkr58E-3	1.24E-07	1.094
85	CG7900	1.31E-07	1.668
86	Pif1A	1.31E-07	2.295
87	PAN2	1.31E-07	1.044
88	SP1029	1.31E-07	0.906
89	CROT	1.31E-07	1.600
90	Drsl2	1.31E-07	7.822
91	ash1	1.38E-07	1.862
92	DNApol-zeta	1.39E-07	1.475
93	CG13694	1.51E-07	2.058
94	cnn	1.51E-07	1.165
95	hob	1.75E-07	1.831
96	CG3552	2.56E-07	1.367
97	CR42530	2.63E-07	2.696
98	Reg-2	3.19E-07	1.438
99	slx1	3.46E-07	2.526
100	CG6686	3.72E-07	1.290

Supplementary Table 3.13: Top 100 upregulated genes with the lowest padj values (T22)

Rank	Gene Symbol	padj value	log2 fc
1	Ugt302E1	1.37E-54	6.009
2	ppk	9.95E-44	6.363
3	lncRNA:CR44536	2.59E-36	8.895
4	Cyp28d2	6.82E-30	3.647
5	Odc2	3.59E-26	4.643
6	CG12643	2.26E-22	3.033
7	asRNA:CR44030	1.05E-20	4.500
8	CG17977	2.98E-18	2.419
9	CG32104	5.80E-18	1.888
10	CR43697	8.95E-17	3.778
11	CG34105	5.55E-15	10.276
12	CG7912	8.08E-15	2.676
13	CG8654	3.53E-14	2.192
14	CG3552	4.34E-14	1.806
15	CG6654	4.38E-14	1.606
16	CG12863	5.04E-14	2.090
17	Cyp6a17	5.04E-14	2.276
18	tut	9.93E-14	3.865
19	CG2064	1.00E-13	2.292
20	Rlip	1.35E-13	1.198
21	CG11601	2.01E-13	1.421
22	lncRNA:CR45039	3.98E-13	4.974
23	CR45470	5.52E-13	4.728
24	Lectin-galC1	1.27E-12	3.439
25	CG12491	2.57E-12	10.696
26	ms(3)76Ba	3.07E-12	9.067
27	CG9344	5.17E-12	2.179
28	CG30059	5.67E-12	3.005
29	CG17177	7.85E-12	5.066
30	CG34215	8.77E-12	2.321
31	Drsl2	8.77E-12	6.204
32	CG4730	1.25E-11	2.640
33	lncRNA:CR42861	1.98E-11	2.182
34	lncRNA:CR44662	1.99E-11	2.906
35	mrt	7.37E-11	1.325
36	CG10353	8.70E-11	1.543
37	CG31098	9.63E-11	1.502
38	CG18278	1.06E-10	2.778
39	CG17803	1.21E-10	1.622
40	CG5612	1.21E-10	2.199
41	CG3939	1.29E-10	1.209
42	twz	1.87E-10	1.925
43	CG10445	2.00E-10	2.686
44	ORMDL	2.36E-10	0.911
45	Fnta	2.86E-10	0.835
46	CG4335	3.97E-10	2.573
47	CG30392	4.49E-10	1.368
48	mRpL54	5.43E-10	1.175
49	CG4627	5.80E-10	1.546
50	CG14882	6.06E-10	1.388
51	FucTC	6.68E-10	2.349
52	Sclp	6.91E-10	1.407
53	CG3397	8.36E-10	2.134
54	narya	8.36E-10	6.139
55	Pif1A	1.10E-09	2.409
56	ZnT35C	1.22E-09	1.592
57	CG10344	1.29E-09	1.590
58	fabp	1.66E-09	1.651
59	CG18858	1.78E-09	1.004
60	CG8001	2.29E-09	1.139
61	mRpL39	2.35E-09	0.988
62	CG7488	2.41E-09	0.902
63	asRNA:CR44960	2.82E-09	3.782
64	CG14711	3.76E-09	0.893
65	PH4alphaPV	3.80E-09	1.406
66	CG42319	4.03E-09	1.863
67	CG33506	4.31E-09	0.991
68	CG6770	4.42E-09	1.257
69	CG5882	4.56E-09	2.670
70	CG13603	4.93E-09	1.219
71	CG33189	4.96E-09	2.254
72	CG43110	6.52E-09	2.246
73	CG6870	6.62E-09	2.266
74	CG12375	7.77E-09	0.859
75	Atg12	8.77E-09	1.386
76	Phae2	9.07E-09	2.044
77	Ssb-c31a	1.02E-08	1.714
78	CG3165	1.07E-08	1.599
79	CG31633	1.14E-08	1.091
80	CR43696	1.23E-08	2.127
81	l(3)mbn	1.33E-08	2.637
82	Paics	1.36E-08	1.187
83	Mics1	1.39E-08	3.856
84	CG33225	1.48E-08	4.269
85	lncRNA:CR44442	1.48E-08	4.272
86	CG17294	1.72E-08	1.542
87	lncRNA:CR43493	1.72E-08	2.175
88	mus201	1.83E-08	1.457
89	lncRNA:CR43242	1.90E-08	2.090
90	CG7054	2.17E-08	1.622
91	CG7900	2.18E-08	1.699
92	Cyp4e3	2.77E-08	2.632
93	CG14499	2.98E-08	3.978
94	CG17568	2.99E-08	1.276
95	CG2246	3.21E-08	0.977
96	niki	3.28E-08	1.645
97	CG34348	3.32E-08	1.425
98	S1P	3.89E-08	1.406
99	CG42822	3.89E-08	2.106
100	CG13339	4.18E-08	1.137

Supplementary Table 3.14: Top 100 upregulated genes with the lowest padj values (T23)

Rank	Gene Symbol	padj value	log2 fc
1	Ugt302E1	1.08E-41	4.851
2	ppk	2.91E-39	6.051
3	lncRNA:CR44536	7.07E-29	9.014
4	Odc2	2.19E-28	4.900
5	Cyp28d2	2.42E-28	3.559
6	CG12643	4.50E-23	3.096
7	CG7912	2.31E-20	3.167
8	mrt	3.26E-20	1.823
9	CG6654	1.47E-16	1.759
10	mRpS9	8.65E-16	0.901
11	CG34105	3.18E-15	10.417
12	twz	4.47E-15	2.342
13	ZnT35C	6.23E-15	2.010
14	lncRNA:CR45039	9.76E-15	4.980
15	asRNA:CR44030	2.42E-14	3.659
16	Fnta	3.81E-14	0.998
17	CR43697	4.21E-14	3.464
18	CG31098	5.95E-14	1.725
19	CG42319	1.14E-13	2.312
20	CG12491	1.45E-13	9.199
21	CG8654	2.54E-13	2.115
22	CG16935	2.68E-13	1.675
23	CG17977	5.26E-13	2.174
24	CG8001	6.76E-13	1.362
25	CG10353	8.42E-13	1.688
26	CG5882	2.45E-12	3.426
27	CG3939	3.51E-12	1.318
28	Rlip	7.68E-12	1.126
29	CG11601	1.19E-11	1.337
30	ORMDL	1.50E-11	0.995
31	Odc1	2.32E-11	2.583
32	PH4alphaPV	3.30E-11	1.579
33	l(3)mbn	3.69E-11	3.128
34	CG1271	9.52E-11	1.372
35	Phae2	1.72E-10	2.252
36	Lectin-galC1	2.13E-10	3.283
37	FucTC	3.19E-10	2.406
38	CG9988	6.51E-10	3.549
39	CG14499	8.30E-10	4.494
40	CG4730	1.41E-09	2.418
41	sni	1.61E-09	1.064
42	CG17177	1.63E-09	5.497
43	CG14711	1.83E-09	0.926
44	CG34215	2.01E-09	2.068
45	CG8541	2.08E-09	3.860
46	GABA-B-R3	2.83E-09	1.361
47	CG2064	2.86E-09	1.869
48	ms(3)76Ba	2.86E-09	8.430
49	CG32104	3.85E-09	1.336
50	CG15356	5.63E-09	1.430
51	betaTub85D	5.63E-09	2.165
52	CG17834	5.72E-09	1.069
53	CG33012	6.19E-09	2.006
54	CG2246	6.38E-09	1.027
55	RabX1	7.47E-09	1.208
56	niki	7.95E-09	1.847
57	CG4335	8.06E-09	2.405
58	CG43110	8.54E-09	2.268
59	lncRNA:CR44442	8.68E-09	5.215
60	axo	1.09E-08	1.550
61	CG17292	1.22E-08	1.151
62	CG44215	1.23E-08	4.838
63	CG34348	1.40E-08	1.472
64	CG12863	1.50E-08	1.642
65	CG6424	1.52E-08	1.018
66	CG18324	2.06E-08	1.371
67	VhaAC39-1	2.08E-08	0.779
68	CG33189	2.14E-08	2.308
69	narya	2.22E-08	4.902
70	mRpL39	2.26E-08	0.940
71	CG32795	2.35E-08	0.969
72	CG8837	2.61E-08	1.761
73	CG4627	2.85E-08	1.451
74	CG33506	3.16E-08	0.965
75	CG13694	3.18E-08	2.123
76	CR43105	3.51E-08	2.369
77	Lerp	4.13E-08	1.079
78	swi2	4.53E-08	1.788
79	CG3520	5.41E-08	1.615
80	Sardh	6.28E-08	1.937
81	CG18278	6.34E-08	2.375
82	CG9147	7.72E-08	1.356
83	Cpr49Ad	8.22E-08	3.737
84	CG18814	8.65E-08	2.760
85	CG17803	8.65E-08	1.375
86	CG15431	8.88E-08	1.626
87	Drsl2	9.07E-08	7.734
88	CG6923	9.27E-08	0.803
89	CG7054	9.55E-08	1.556
90	tut	1.00E-07	2.805
91	CG32335	1.02E-07	1.770
92	beta4GalNAcTB	1.07E-07	1.265
93	CG15044	1.37E-07	1.205
94	PH4alphaNE1	1.38E-07	1.975
95	CG5612	1.41E-07	1.834
96	CG7900	1.42E-07	1.615
97	Lpin	1.53E-07	1.208
98	CG43191	1.65E-07	3.188
99	CR43696	1.84E-07	2.047
100	Fpgs	2.22E-07	1.400

General concluding perspective

Within this thesis project, I set out to characterise the functional roles of lncRNAs using *Drosophila* melanogaster as the model organism of choice. Even though lncRNAs were found to be involved in various biological processes, their functional relevance was disregarded for a long time. Here, I have presented the approach that I have taken to study lncRNA function during *Drosophila* embryogenesis and demonstrated that lncRNAs could be studied in vivo. Using RAMPAGE data, I have first selected for ncRNAs that fulfilled the definition of a lncRNA transcript and conserved expression profile across 5 *Drosophila* species. From this, I have identified those that were expressed during *Drosophila* embryonic development and selected 22 of them to determine their functional relevance.

Using genome editing tools such as CRISPR, I have removed the TSS of our lncRNA candidates to generate lncRNA mutants for our study. To determine if they have any function, I have created a simple set of tests to identify the phenotype of the mutation. Of the candidates lncRNAs that were successfully generated, I have identified two candidate lncRNAs that were homozygous lethal. This number is impressive given that we have first selected for our lncRNA candidates using a strict filter and the chance of generating a lncRNA with a lethal phenotype was 2 in 21. Given this ratio, I would not be surprise if there were more lncRNAs in the *Drosophila* genome with important functions and the deletion of these transcripts would be lethal for the fruit fly. That said, as this study was carried out as a pilot project with a small number of lncRNA candidates, more large-scale genome editing projects would be needed to investigate the number of functional lncRNAs transcripts in the genome given the wealth of data that we have about *Drosophila* today.

As lncRNAs could be involved in a myriad of biological processes, I have designed a number of assays to study the molecular phenotype of the lncRNA knockout. Majority of these mutants had no obvious anatomical or functional abnormalities such as sterility or locomotion issues. This result coincides with other lncRNA-related studies that were published recently to understand the role of lncRNA during embryogenesis (Ignacio *et al.,* 2018; Goudarzi *et al.,* 2019). The reason for the lack of functional readout from the mutants could be multi-fold such as redundancy. However, majority of these functional assays including ours were centred around the role of lncRNAs in viability and fertility and little was done to further characterise them on a behavioural level given that many lncRNAs were found

to be highly expressed in the nervous system and might have a cognitive function (Briggs *et al.,* 2015; Molyneaux *et al.,* 2015). Therefore, it would be necessary to include more behavioural assays when screening lncRNA mutants. To improve on the experimental design, using a set of lncRNAs that are expressed in a particular tissue or cell type would allow easier planning of phenotypic studies. In a study by McCorkindale *et al.,* 2019, the authors focused on critical events that took place during *Drosophila* embryogenesis and performed RNA-seq on various types of neurogenic cells that were purified across consecutive time points. The data not only identified conserved lncRNAs that are expressed during Drosophila neurogenesis but more importantly, the lncRNAs that that expressed in each cell type is known. Using this information, different phenotypic studies can be carried to understand the impact of neurogenic lncRNAs loss of function in shaping the nervous system development.

From the phenotypic study that was performed, I have shown that it is relevant to invest in time to study lncRNA function in general. Although the 2 lncRNA candidates were homozygous lethality, further validation will be needed to authenticate the molecular phenoptype for lncRNA-3. During the characterisation of the lncRNA candidate lncRNA-9, it was revealed that the lncRNA has a role in the central nervous system. While the GO enrichment analysis has pointed to activities related to chemical synaptic transmission, further mechanistic studies would be needed to understand how the lncRNA candidate interreacted with different biochemical partners for normal physiological behaviours.

Finally, I have observed a shift away from a protein-centric view in modern biology to encompass the roles played by different ncRNAs in gene regulation today. It would be helpful to have a better understanding of lncRNA biology as many of them have been overlooked. The advert of high throughout sequencing technology has unrevealed thousands of lncRNA transcripts that play diverse scanarios. Various studies have focused on identifying lncRNAs that are expressed in specific subcellular domains (nuclear and cytoplasmic) or involved in a biological process such as dosage compensation (Landskron *et al.,* 2018; Meller *et al.,* 1997). Several studies, including ours, have identified lncRNAs that are evident in the *Drosophila* nervous system (Schor *et al.,* 2018; McCorkindale *et al.,* 2019). The challenging task remains to unravel the molecular mechanisms by which they act. Further on from this study, it will be valuable to investigate how lncRNA-9 and many others that are expressed in the nervous system contributes to *Drosophila* neuronal functions.

Bibliography

Aboobaker, A. A., Tomancak, P., Patel, N., Rubin, G. M. & Lai, E. C. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc. Natl. Acad. Sci. U.S.A. 102, 18017–18022 (2005).

Adams, M. D. et al. The genome sequence of Drosophila melanogaster. Science 287, 2185–2195 (2000).

Akay, A. et al. Identification of functional long non-coding RNAs in C. elegans. BMC Biology 17, 14 (2019).

Alderson, T. Chemically induced delayed germinal mutation in Drosophila. Nature 207, 164–167 (1965).

Amaral, P. P., Clark, M. B., Gascoigne, D. K., Dinger, M. E. & Mattick, J. S. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 39, D146-151 (2011).

Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. Mamm. Genome 19, 454–492 (2008).

Andergassen, D. et al. The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. PLoS Genet 15, (2019).

Andersen, R. E. & Lim, D. A. Forging our understanding of lncRNAs in the brain. Cell Tissue Res. 371, 55–71 (2018).

Andrews, S. J. & Rothnagel, J. A. Emerging evidence for functional peptides encoded by short open reading frames. Nature Reviews Genetics 15, 286 (2014).

Ang, C. E. et al. The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. eLife https://elifesciences.org/articles/41770 (2019) doi:10.7554/eLife.41770.

Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761–764 (2007).

Ardehali, M. B. et al. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J 30, 2817–2828 (2011).

Arun, G. et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. Genes Dev. 30, 34–51 (2016).

Atkinson, S. R. et al. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. RNA 24, 1195–1213 (2018).

Baker, K. D. & Thummel, C. S. Diabetic larvae and obese flies-emerging studies of metabolism in Drosophila. Cell Metab. 6, 257–266 (2007).

Baldassarre, A. & Masotti, A. Long Non-Coding RNAs and p53 Regulation. Int J Mol Sci 13, 16708–16717 (2012).

Ballabio, A. & Willard, H. F. Mammalian X-chromosome inactivation and the XIST gene. Curr. Opin. Genet. Dev. 2, 439–447 (1992).

Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113 (2004).

Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297 (2004).

Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. Cell Rep 4, 220–228 (2013).

Bateman, J. R., Lee, A. M. & Wu, C. -ting. Site-Specific Transformation of Drosophila via ϕC31 Integrase-Mediated Cassette Exchange. Genetics 173, 769–777 (2006).

Batut, P., Dobin, A., Plessy, C., Carninci, P. & Gingeras, T. R. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 23, 169–180 (2013).

Batut, P. & Gingeras, T. R. Rampage: Promoter Activity Profiling by Paired-end Sequencing of 5′-complete cDNAs. Curr Protoc Mol Biol 104, Unit-25B.11 (2013).

Batut, P. J. & Gingeras, T. R. Conserved noncoding transcription and core promoter regulatory code in early Drosophila development. eLife 6,.

Baulcombe, D. Overview of RNA Interference and Related Processes. Current Protocols in Molecular Biology 62, 26.1.1-26.1.6 (2003).

Bazzini, A. A. et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. 33, 981–993 (2014).

Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A 27, 499–506 (1941).

Beira, J. V. & Paro, R. The legacy of Drosophila imaginal discs. Chromosoma 125, 573–592 (2016).

Bejarano, F., Smibert, P. & Lai, E. C. miR-9a prevents apoptosis during wing development by repressing Drosophila LIM-only. Dev. Biol. 338, 63–73 (2010).

Bendena, W. G., Ayme-Southgate, A., Garbe, J. C. & Pardue, M. L. Expression of heat-shock locus hsr-omega in nonstressed cells during development in Drosophila melanogaster. Developmental Biology 144, 65–77 (1991).

Bernard, D. et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J 29, 3082–3093 (2010).

Bernstein, E. et al. Dicer is essential for mouse development. Nat. Genet. 35, 215–217 (2003).

Beumer, K. J. et al. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. U.S.A. 105, 19821–19826 (2008).

Bhadra, U., Pal-Bhadra, M. & Birchler, J. A. Role of the male specific lethal (msl) Genes in Modifying the Effects of Sex Chromosomal Dosage in Drosophila. Genetics 152, 249–268 (1999).

Bhat, S. A. et al. Long non-coding RNAs: Mechanism of action and functional utility. Noncoding RNA Res 1, 43–50 (2016).

Bhatia, G., Sharma, S., Upadhyay, S. K. & Singh, K. Long Non-coding RNAs Coordinate Developmental Transitions and Other Key Biological Processes in Grapevine. Sci Rep 9, (2019).

Bier, E. Drosophila, the golden bug, emerges as a tool for human genetics. Nat. Rev. Genet. 6, 9–23 (2005).

Bier, E., Harrison, M. M., O'Connor-Giles, K. M. & Wildonger, J. Advances in Engineering the Fly Genome with the CRISPR-Cas System. Genetics 208, 1–18 (2018).

Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. Nature 422, 633–637 (2003).

Blumenstiel, J. P. et al. Identification of EMS-induced mutations in Drosophila melanogaster by whole-genome sequencing. Genetics 182, 25–32 (2009).

Boettcher, M. & McManus, M. T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. Mol. Cell 58, 575–585 (2015).

Boley, N., Wan, K. H., Bickel, P. J. & Celniker, S. E. Navigating and Mining modENCODE Data. Methods 68, 38–47 (2014).

Bond, A. M. et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nature Neuroscience 12, 1020–1027 (2009).

Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. 44, 148–156 (2012).

Borsani, G. et al. Characterization of a murine gene expressed from the inactive X chromosome. Nature 351, 325–329 (1991).

Böttcher, R. et al. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res. 42, e89 (2014).

Bownes, M. The Embryonic Development of Drosophila Melanogaster. Jose A. Campos-Ortega , Volker Hartenstein. The Quarterly Review of Biology 61, 404–404 (1986).

Braidotti, G. et al. The Air Noncoding RNA: An Imprinted cis-silencing Transcript. Cold Spring Harb Symp Quant Biol 69, 55–66 (2004).

Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401–415 (1993).

Brennecke, J. et al. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322, 1387–1392 (2008).

Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L. & Barry, G. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. Neuron 88, 861–877 (2015).

Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol 46, 111–138 (1971).

Britten, R. J. & Davidson, E. H. Gene Regulation for Higher Cells: A Theory. Science 165, 349–357 (1969).

Broadbent, H. M. et al. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum. Mol. Genet. 17, 806–814 (2008).

Brockdorff, N. et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526 (1992).

Brown, C. J. et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71, 527–542 (1992).

Brown, J. B. et al. Diversity and dynamics of the Drosophila transcriptome. Nature 512, 393–399 (2014).

Brown, J. B. & Celniker, S. E. Lessons from modENCODE. Annu Rev Genomics Hum Genet 16, 31–53 (2015).

Burge, C. et al. Splicing of precursors to mRNAs by the spliceosomes. (1999).

Cabezas-Wallscheid, N. et al. Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. Cell Stem Cell 15, 507–522 (2014).

Cabili, M. N. et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biology 16, 20 (2015).

Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927 (2011).

Cáceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L. & Krainer, A. R. Role of the Modular Domains of SR Proteins in Subnuclear Localization and Alternative Splicing Specificity. J Cell Biol 138, 225–238 (1997).

Cantarella, C. et al. A next generation sequencing-based approach to identify piRNAs in

breast cancer cells. EMBnet.journal 18, 96–97 (2012).
Carmona, S., Lin, B., Chou, T., Arroyo, K. & Sun, S. LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. PLoS Genet. 14, e1007378 (2018).

Carninci, P. et al. The transcriptional landscape of the mammalian genome. Science 309, 1559–1563 (2005).

Carninci, P. et al. Targeting a Complex Transcriptome: The Construction of the Mouse Full-Length cDNA Encyclopedia. Genome Res 13, 1273–1289 (2003).

Carrieri, C. et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491, 454–457 (2012).

Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642–655 (2009).

Cesana, M. et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369 (2011).

Chalei, V. et al. The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. eLife 3,.

Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. Nat. Rev. Genet. 8, 884–896 (2007).

Chawla, G. & Sokol, N. S. MicroRNAs in Drosophila development. Int Rev Cell Mol Biol 286, 1–65 (2011).

Chen, B. et al. Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. Sci Rep 6, 23330 (2016).

Chen, F., Ji, J., Shen, J. & Lu, X. When Long Noncoding RNAs Meet Genome Editing in Pluripotent Stem Cells. Stem Cells Int 2017, (2017).

Chen, G. et al. LncRNA SRA promotes hepatic steatosis through repressing the expression of adipose triglyceride lipase (ATGL). Sci Rep 6, (2016).

Chen, J. C. et al. Morpholino-mediated Knockdown of DUX4 Toward Facioscapulohumeral Muscular Dystrophy Therapeutics. Mol Ther 24, 1405–1411 (2016).

Chen, M.-J. M. et al. Integrating RNA-seq and ChIP-seq data to characterize long non-coding RNAs in Drosophila melanogaster. BMC Genomics 17, 220 (2016).

Chen, X. et al. AlncRNA HULC as an effective biomarker for surveillance of the outcome of cancer: A meta-analysis. PLoS One 12, (2017).

Chen, Y.-W. et al. Systematic study of Drosophila microRNA functions using a collection of targeted knockout mutations. Dev. Cell 31, 784–800 (2014).

Chen, Z.-X. et al. Comparative validation of the D. melanogaster modENCODE transcriptome annotation. Genome Res. 24, 1209–1223 (2014).

Chew, G.-L. et al. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Development 140, 2828–2834 (2013).

Chintapalli, V. R., Al Bratty, M., Korzekwa, D., Watson, D. G. & Dow, J. A. T. Mapping an Atlas of Tissue-Specific Drosophila melanogaster Metabolomes by High Resolution Mass Spectrometry. PLoS One 8, (2013).

Chintapalli, V. R., Wang, J. & Dow, J. A. T. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat. Genet. 39, 715–720 (2007).

Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 31, 230–232 (2013).

Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145, dev165753 (2018).

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell 44, 667–678 (2011).

Chu, C., Spitale, R. C. & Chang, H. Y. Technologies to probe functions and mechanisms of long noncoding RNAs. Nat. Struct. Mol. Biol. 22, 29–35 (2015).

Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823 (2013).

Cooper, C. et al. Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. Nucleic Acids Res 37, 4518–4531 (2009).

Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. 41, 9584–9592 (2013).

Crick, F. On protein synthesis.

Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. Nat. Rev. Genet. 16, 57–66 (2015).

Czech, B. et al. An endogenous small interfering RNA pathway in Drosophila. Nature 453, 798–802 (2008).

Czech, B. et al. piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annual Review of Genetics 52, 131–157 (2018).

Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol. Cell 50, 749–761 (2013).

Dahl, M., Kristensen, L. S. & Grønbæk, K. Long Non-Coding RNAs Guide the Fine-Tuning

of Gene Regulation in B-Cell Development and Malignancy. International Journal of Molecular Sciences 19, 2475 (2018).

Dai, Q., Smibert, P. & Lai, E. C. Exploiting Drosophila genetics to understand microRNA function and regulation. Curr Top Dev Biol 99, 201–235 (2012).

Davidovich, C. & Cech, T. R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA 21, 2007–2022 (2015).

Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22, 1775–1789 (2012).

DiCarlo, J. E. et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 41, 4336–4343 (2013).

Dieci, G., Preti, M. & Montanini, B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics 94, 83–88 (2009).

Diederichs, S. The four dimensions of noncoding RNA conservation. Trends Genet. 30, 121–123 (2014).

Dimitrova, N. et al. LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. Mol Cell 54, 777–790 (2014).

Dinger, M. E., Pang, K. C., Mercer, T. R. & Mattick, J. S. Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. PLoS Comput Biol 4, (2008).

Dittmar, K. A., Goodenbour, J. M. & Pan, T. Tissue-specific differences in human transfer RNA expression. PLoS Genet. 2, e221 (2006).

Djebali, S. et al. Landscape of transcription in human cells. Nature 489, 101–108 (2012).

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).

Dreos, R., Ambrosini, G., Cavin Périer, R. & Bucher, P. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. Nucleic Acids Res. 41, D157-164 (2013).

Droujinine, I. A. & Perrimon, N. Defining the interorgan communication network: systemic coordination of organismal cellular processes under homeostasis and localized stress. Front Cell Infect Microbiol 3, (2013).

Dunagin, M., Cabili, M. N., Rinn, J. & Raj, A. Visualization of lncRNA by single-molecule fluorescence in situ hybridization. Methods Mol. Biol. 1262, 3–19 (2015).

Dye, N. A. et al. Cell dynamics underlying oriented growth of the Drosophila wing imaginal disc. Development 144, 4406–4421 (2017).

Eißmann, M. et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol 9, 1076–1087 (2012).

ENCODE Project Consortium et al. Identification and analysis of functional elements in 1%

of the human genome by the ENCODE pilot project. Nature 447, 799–816 (2007).

Engreitz, J., Lander, E. S. & Guttman, M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Methods Mol. Biol. 1262, 183–197 (2015).

Eoh, K. J. et al. Long non-coding RNA, steroid receptor RNA activator (SRA), induces tumor proliferation and invasion through the NOTCH pathway in cervical cancer cell lines. Oncol Rep 38, 3481–3488 (2017).

Epler, J. L. Ethyl Methanesulfonate-Induced Lethals in Drosophila—frequency-Dose Relations and Multiple Mosaicism. Genetics 54, 31–36 (1966).

Erard, N., Knott, S. R. V. & Hannon, G. J. A CRISPR Resource for Individual, Combinatorial, or Multiplexed Gene Knockout. Mol. Cell 67, 348-354.e4 (2017).

Essletzbichler, P. et al. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. 24, 2059–2065 (2014).

Fagegaltier, D. et al. Oncogenic transformation of Drosophila somatic cells induces a functional piRNA pathway. Genes Dev. 30, 1623–1635 (2016).

Faghihi, M. A. et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol. 11, R56 (2010).

Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806–811 (1998).

Flynn, R. A. et al. Dissecting noncoding and pathogen RNA-protein interactomes. RNA 21, 135–143 (2015).

Franke, A. & Baker, B. S. The rox1 and rox2 RNAs Are Essential Components of the Compensasome, which Mediates Dosage Compensation in Drosophila. Molecular Cell 4, 117–122 (1999).

Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. Nature 171, 740 (1953).

Frenkel, L. et al. Organization of Circadian Behavior Relies on Glycinergic Transmission. Cell Rep 19, 72–85 (2017).

Froberg, J. E., Yang, L. & Lee, J. T. Guided by RNAs: X-inactivation as a model for lncRNA function. J. Mol. Biol. 425, 3698–3706 (2013).

Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31, 397–405 (2013).

Ge, D. T., Tipping, C., Brodsky, M. H. & Zamore, P. D. Rapid Screening for CRISPR-Directed Editing of the Drosophila Genome Using white Coconversion. G3 (Bethesda) 6, 3197–3206 (2016).

Gebert, L. F. R. & MacRae, I. J. Regulation of microRNA function in animals. Nat. Rev. Mol. Cell Biol. 20, 21–37 (2019).

Georgiev, P., Chlamydas, S. & Akhtar, A. Drosophila dosage compensation. Fly (Austin) 5, 147–154 (2011).

Geslain, R. & Pan, T. Functional analysis of human tRNA isodecoders. J. Mol. Biol. 396, 821–831 (2010).

Giambruno, R., Mihailovich, M. & Bonaldi, T. Mass Spectrometry-Based Proteomics to Unveil the Non-coding RNA World. Front. Mol. Biosci. 5, (2018).

Giraldez, A. J. et al. MicroRNAs regulate brain morphogenesis in zebrafish. Science 308, 833–838 (2005).

Gitaí, D. L. G. et al. The non-coding RNA BC1 is down-regulated in the hippocampus of Wistar Audiogenic Rat (WAR) strain after audiogenic kindling. Brain Res. 1367, 114–121 (2011).

Glaser, A., McColl, B. & Vadolas, J. GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. Mol Ther Nucleic Acids 5, e334 (2016).

Goff, L. A. & Rinn, J. L. Linking RNA biology to lncRNAs. Genome Res. 25, 1456–1465 (2015).

Gomes, A. Q., Nolasco, S. & Soares, H. Non-coding RNAs: multi-tasking molecules in the cell. Int J Mol Sci 14, 16010–16039 (2013).

Gong, C., Tang, Y. & Maquat, L. E. mRNA–mRNA duplexes that auto-elicit Staufen1-mediated mRNA decay. Nat Struct Mol Biol 20, 1214–1220 (2013).

Gonzalez, J., Qi, H., Liu, N. & Lin, H. Piwi Is a Key Regulator of Both Somatic and Germline Stem Cells in the Drosophila Testis. Cell Reports 12, 150–161 (2015).

Goudarzi, M., Berg, K., Pieper, L. M. & Schier, A. F. Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. Elife 8, (2019).

Gratz, S. J., Wildonger, J., Harrison, M. M. & O'Connor-Giles, K. M. CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. Fly (Austin) 7, 249–255 (2013).

Gratz, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196, 961–971 (2014).

Grote, P. et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell 24, 206–214 (2013).

Gunawardane, L. S. et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science 315, 1587–1590 (2007).

Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076 (2010).

Gutschner, T. et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res. 73, 1180–1189 (2013).

Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227 (2009).

Guttman, M. et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300 (2011).

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large non-coding RNAs do not encode proteins. Cell 154, 240–251 (2013).

Halder, G. & Mills, G. B. Drosophila in cancer research: to boldly go where no one has gone before. Oncogene 30, 4063–4066 (2011).

Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950–952 (1999).

Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. PLOS Genetics 9, e1003569 (2013).

Hannon, G. J. RNA interference. Nature 418, 244–251 (2002).

Hardiman, K. E., Brewster, R., Khan, S. M., Deo, M. & Bodmer, R. The bereft gene, a potential target of the neural selector gene cut, contributes to bristle morphogenesis. Genetics 161, 231–247 (2002).

Harris Harry, Watts J. W. & Abraham Edward Penley. The relationship between nuclear and cytoplasmic ribonucleic acid. Proceedings of the Royal Society of London. Series B. Biological Sciences 156, 109–121 (1962).

Hayashi, I., Oyama, T. & Morikawa, K. Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus. EMBO J. 20, 1819–1828 (2001).

He, R.-Z., Luo, D.-X. & Mo, Y.-Y. Emerging roles of lncRNAs in the post-transcriptional regulation in cancer. Genes Dis 6, 6–15 (2019).

Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. Nat. Methods 11, 122–123 (2014).

Heo, J. B., Lee, Y.-S. & Sung, S. Epigenetic regulation by long noncoding RNAs in plants. Chromosome Res 21, 685–693 (2013).

Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 331, 76–79 (2011).

Heo, J. B. & Sung, S. Encoding memory of winter by noncoding RNAs. Epigenetics 6,

544–547 (2011).

Heraud-Farlow, J. E. & Kiebler, M. A. The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. Trends Neurosci. 37, 470–479 (2014).

Heraud-Farlow, J. E. et al. Staufen2 regulates neuronal target RNAs. Cell Rep 5, 1511–1518 (2013).

Herriges, M. J. et al. Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. Genes Dev. 28, 1363–1379 (2014).

Holley, R. W. et al. STRUCTURE OF A RIBONUCLEIC ACID. Science 147, 1462–1465 (1965).

Hon, C.-C. et al. An atlas of human long non-coding RNAs with accurate 5′ ends. Nature 543, 199–204 (2017).

Housden, B. E., Lin, S. & Perrimon, N. Cas9-based genome editing in Drosophila. Meth. Enzymol. 546, 415–439 (2014).

Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832 (2013).

Hu, G. et al. Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis. Oncotarget 9, 18648–18663 (2018).

Huang, A. M., Rehm, E. J. & Rubin, G. M. Recovery of DNA sequences flanking P-element insertions in Drosophila: inverse PCR and plasmid rescue. Cold Spring Harb Protoc 2009, pdb.prot5199 (2009).

Huang, J.-Z. et al. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. Mol. Cell 68, 171-184.e6 (2017).

Huang, Y., Gu, L. & Li, G.-M. H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. J. Biol. Chem. 293, 7811–7823 (2018).
174.Huarte, M. The emerging role of lncRNAs in cancer. Nature Medicine 21, 1253–1261 (2015).

Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142, 409–419 (2010).

Hung, T. et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat. Genet. 43, 621–629 (2011).

Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–229 (2013).

Ilik, I. & Akhtar, A. roX RNAs: non-coding regulators of the male X chromosome in flies. RNA Biol 6, 113–121 (2009).

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome

profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols 7, 1534–1550 (2012).

Iorio, M. V. et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 65, 7065–7070 (2005).

Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. 47, 199–208 (2015).

Jandura, A., Hu, J., Wilk, R. & Krause, H. M. High Resolution Fluorescent In Situ Hybridization in Drosophila Embryos and Tissues Using Tyramide Signal Amplification. J Vis Exp (2017) doi:10.3791/56281.

Ji, P. et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22, 8031–8041 (2003).

Jiang, W., Liu, Y., Liu, R., Zhang, K. & Zhang, Y. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. Cell Rep 11, 137–148 (2015).

Jin, K. et al. Long non-coding RNA PVT1 interacts with MYC and its downstream molecules to synergistically promote tumorigenesis. Cell. Mol. Life Sci. 76, 4275–4289 (2019).

Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).

Johansson, J. et al. An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell 110, 551–561 (2002).

Jolly, C. & Lakhotia, S. C. Human sat III and Drosophila hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. Nucleic Acids Res. 34, 5508–5514 (2006).

Kallen, A. N. et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol. Cell 52, 101–112 (2013).

Kapranov, P. et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484–1488 (2007).

Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8, 413–423 (2007).

Kapusta, A. & Feschotte, C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends Genet 30, 439–452 (2014).

Katayama, S. et al. Antisense transcription in the mammalian transcriptome. Science 309, 1564–1566 (2005).

Kaushik, K. et al. Dynamic Expression of Long Non-Coding RNAs (lncRNAs) in Adult Zebrafish. PLOS ONE 8, e83616 (2013).

Kelley, R. L. et al. Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell 98, 513–522 (1999).

Kellis, M. et al. Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. U.S.A. 111, 6131–6138 (2014).

Ketting, R. F. et al. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev 15, 2654–2659 (2001).

Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U.S.A. 106, 11667–11672 (2009).

Khandelwal, A., Bacolla, A., Vasquez, K. M. & Jain, A. Long non-coding RNA: A new paradigm for lung cancer. Mol. Carcinog. 54, 1235–1251 (2015).

Khandelwal, A., Malhotra, A., Jain, M., Vasquez, K. M. & Jain, A. The emerging role of long non-coding RNA in gallbladder cancer pathogenesis. Biochimie 132, 152–160 (2017).

Khyzha, N. et al. Regulation of CCL2 expression in human vascular endothelial cells by a neighboring divergently transcribed long noncoding RNA. Proc. Natl. Acad. Sci. U.S.A. 116, 16410–16419 (2019).

Kim, J. et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nat. Genet. 50, 1705–1715 (2018).

Kim, Y. K., Furic, L., Desgroseillers, L. & Maquat, L. E. Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. Cell 120, 195–208 (2005).

Kim, Y. K. et al. Staufen1 regulates diverse classes of mammalian transcripts. EMBO J 26, 2670–2681 (2007).

Kimelman, D. & Martin, B. L. Anterior-Posterior Patterning in Early Development: Three Strategies. Wiley Interdiscip Rev Dev Biol 1, 253–266 (2012).

Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 3, ra8 (2010).

Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell 109, 145–148 (2002).

Klattenhoff, C. A. et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152, 570–583 (2013).

Kloosterman, W. P. & Plasterk, R. H. A. The diverse functions of microRNAs in animal development and disease. Dev. Cell 11, 441–450 (2006).

Koenecke, N., Johnston, J., He, Q., Meier, S. & Zeitlinger, J. Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res 27, 64–74 (2017).

Kok, F. O., Gupta, A., Lawson, N. D. & Wolfe, S. A. Construction and application of site-specific artificial nucleases for targeted gene editing. Methods Mol. Biol. 1101, 267–303 (2014).

Kondo, S. & Ueda, R. Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in Drosophila. Genetics 195, 715–721 (2013).

König, J., Zarnack, K., Luscombe, N. M. & Ule, J. Protein-RNA interactions: new genomic technologies and perspectives. Nat. Rev. Genet. 13, 77–83 (2012).

Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell 172, 393–407 (2018).

Korostowski, L., Sedlak, N. & Engel, N. The Kcnq1ot1 Long Non-Coding RNA Affects Chromatin Conformation and Expression of Kcnq1, but Does Not Regulate Its Imprinting in the Developing Heart. PLOS Genetics 8, e1002956 (2012).

Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 47, D155–D162 (2019).

Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152-157 (2011).

Kretz, M. TINCR, staufen1, and cellular differentiation. RNA Biol 10, 1597–1601 (2013).

Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. Genetics 193, 651–669 (2013).

Kurihara, M. et al. A Testis-Specific Long Non-Coding RNA, lncRNA-Tcam1, Regulates Immune-Related Genes in Mouse Male Germ Cells. Front Endocrinol (Lausanne) 8, (2017).

Lai, K., Jia, S., Yu, S., Luo, J. & He, Y. Genome-wide analysis of aberrantly expressed lncRNAs and miRNAs with associated co-expression and ceRNA networks in β-thalassemia and hereditary persistence of fetal hemoglobin. Oncotarget 8, 49931–49943 (2017).

Lakhotia, S. C., Mallik, M., Singh, A. K. & Ray, M. The large noncoding hsrω-n transcripts are essential for thermotolerance and remobilization of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in Drosophila. Chromosoma 121, 49–70 (2012).

Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).

Landskron, L. et al. The asymmetrically segregating lncRNA cherub is required for transforming stem cells into malignant cells. Elife 7, (2018).

Langridge, R. et al. The molecular configuration of deoxyribonucleic acid: II. Molecular models and their fourier transforms. Journal of Molecular Biology 2, 38-IN12 (1960).

Lanz, R. B. et al. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell 97, 17–27 (1999).

Latos, P. A. et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338, 1469–1472 (2012).

Lee, B. et al. Long Noncoding RNAs as Putative Biomarkers for Prostate Cancer Detection. J Mol Diagn 16, 615–626 (2014).

Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. 21, 400–404 (1999).

Lee, T. I. & Young, R. A. Transcriptional Regulation and its Misregulation in Disease. Cell 152, 1237–1251 (2013).

Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature 375, 34–39 (1995).

Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. Nucleic Acids Res. 44, 863–877 (2016).

Leone, S. & Santoro, R. Challenges in the analysis of long noncoding RNA functionality. FEBS Lett. 590, 2342–2353 (2016).

Levsky, J. M. & Singer, R. H. Fluorescence in situ hybridization: past, present and future. Journal of Cell Science 116, 2833–2838 (2003).

Lewis, E. B. A gene complex controlling segmentation in Drosophila. Nature 276, 565–570 (1978).

Li, L. & Chang, H. Y. Physiological roles of long noncoding RNAs: insight from knockout mice. Trends Cell Biol. 24, 594–602 (2014).

Li, L. et al. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell Rep 5, 3–12 (2013).

Li, M. et al. The novel long non-coding RNA CRG regulates Drosophila locomotor behavior. Nucleic Acids Res 40, 11714–11727 (2012).

Li, Z., Zhao, L. & Wang, Q. Overexpression of long non-coding RNA HOTTIP increases chemoresistance of osteosarcoma cell by activating the Wnt/β-catenin pathway. Am J Transl Res 8, 2385–2393 (2016).

Liang, M. et al. Sequential expression of long noncoding RNA as mRNA gene expression in specific stages of mouse spermatogenesis. Sci Rep 4, (2014).

Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. J. Biotechnol. 241, 136–146 (2017).

Lin, N. et al. An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. Molecular Cell 53, 1005–1019 (2014).

Lin, S., Ewen-Campen, B., Ni, X., Housden, B. E. & Perrimon, N. In Vivo Transcriptional Activation Using CRISPR/Cas9 in Drosophila. Genetics 201, 433–442 (2015).

Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. Nature 431, 364–370 (2004).

Lipshitz, H. D., Peattie, D. A. & Hogness, D. S. Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes Dev. 1, 307–322 (1987).

Liu, B. et al. lncRNA GAS5 Reverses EMT and Tumor Stem Cell-Mediated Gemcitabine Resistance and Metastasis by Targeting miR-221/SOCS3 in Pancreatic Cancer. Mol Ther Nucleic Acids 13, 472–482 (2018).

Liu, J. et al. Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics 39, 209–215 (2012).

Liu, S. J. et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 355, (2017).

Liu, X.-H. et al. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. Mol. Cancer 13, 92 (2014).

Louro, R., Smirnova, A. S. & Verjovski-Almeida, S. Long intronic noncoding RNA transcription: expression noise or expression choice? Genomics 93, 291–298 (2009).

Lu, B. & Vogel, H. Drosophila models of neurodegenerative diseases. Annu Rev Pathol 4, 315–342 (2009).

Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature 555, 107–111 (2018).

Ma, C. et al. The growth arrest-specific transcript 5 (GAS5): a pivotal tumor suppressor long noncoding RNA in human cancers. Tumour Biol. 37, 1437–1444 (2016).

Majumdar, S. & Rio, D. C. P Transposable Elements in Drosophila and other Eukaryotic Organisms. Microbiol Spectr 3, MDNA3-0004–2014 (2015).

Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).

Mallik, M. & Lakhotia, S. C. RNAi for the large non-coding hsromega transcripts suppresses polyglutamine pathogenesis in Drosophila models. RNA Biol 6, 464–478 (2009).

Mancini-DiNardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev 20, 1268–1282 (2006).

Marahrens, Y., Loring, J. & Jaenisch, R. Role of the Xist gene in X chromosome choosing. Cell 92, 657–664 (1998).

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11, 156–166 (1997).

Matsumoto, A. et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature 541, 228–232 (2017).
Matthews, B. B. et al. Gene Model Annotations for Drosophila melanogaster: Impact of High-Throughput Data. G3: Genes, Genomes, Genetics 5, 1721–1736 (2015).

Mattick, J. S. RNA regulation: a new genetics? Nat. Rev. Genet. 5, 316–323 (2004).

Mattick, J. S. Probing the phenomics of noncoding RNA. Elife 2, e01968 (2013).

Mattick, J. S. & Makunin, I. V. Non-coding RNA. Hum. Mol. Genet. 15 Spec No 1, R17-29 (2006).

McCorkindale, A. L. et al. A gene expression atlas of embryonic neurogenesis in Drosophila reveals complex spatiotemporal regulation of lncRNAs. Development 146, (2019).

Mehtab, M. Chromosomal re-arrangements in the progeny of Drosophila males treated with mustard gas. Nature 171, 262 (1953).

Meier, J. A., Zhang, F. & Sanjana, N. E. GUIDES: sgRNA design for loss-of-function screens. Nat Methods 14, 831–832 (2017).

Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. & Davis, R. L. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell 88, 445–457 (1997).

Meller, V. H. & Rattner, B. P. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. 21, 1084–1091 (2002).

Minton, K. Immune regulation: Long non-coding RNAs in the immune system. Nature Reviews Immunology 13, 617 (2013).

Miyagawa, R. et al. Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. RNA 18, 738–751 (2012).

modENCODE Consortium et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787–1797 (2010).

Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K. & Kanduri, C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development 137, 2493–2499 (2010).

Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K. & Kanduri, C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development 137, 2493–2499 (2010).

Moran, V. A., Perera, R. J. & Khalil, A. M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res 40, 6391–6400 (2012).

Morante, J., Vallejo, D. M., Desplan, C. & Dominguez, M. Conserved miR-8/miR-200

defines a glial niche that controls neuroepithelial expansion and neuroblast transition. Dev. Cell 27, 174–187 (2013).

Morceau, F., Chateauvieux, S., Gaigneaux, A., Dicato, M. & Diederich, M. Long and Short Non-Coding RNAs as Regulators of Hematopoietic Differentiation. Int J Mol Sci 14, 14744–14770 (2013).

Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. Nat. Rev. Genet. 15, 423–437 (2014).

Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V. L., Farzaneh, F. & Williams, G. T. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene 28, 195–208 (2009).

Mozzetta, C. et al. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol. Cell 53, 277–289 (2014).

Muddashetty, R. et al. Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. J. Mol. Biol. 321, 433–445 (2002).

Murchison, E. P. et al. Critical roles for Dicer in the female germline. Genes Dev. 21, 682–693 (2007).

Nagano, T. et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322, 1717–1720 (2008).

Nakagawa, S. et al. Malat1 is not an essential component of nuclear speckles in mice. RNA 18, 1487–1499 (2012).

Nakaya, H. I. et al. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. Genome Biol. 8, R43 (2007).

Nam, J.-W. & Bartel, D. P. Long noncoding RNAs in C. elegans. Genome Res. 22, 2529–2540 (2012).

Nam, J.-W., Choi, S.-W. & You, B.-H. Incredible RNA: Dual Functions of Coding and Noncoding. Mol Cells 39, 367–374 (2016).

Nasrat, G. E., Kaplan, W. D. & Auerbach, C. A quantitative study of mustard gas induced chromosome breaks and re-arrangements in Drosophila Melanogaster. Z Indukt Abstamm Vererbungsl 86, 249–262 (1954).

Ng, S.-Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. Mol. Cell 51, 349–359 (2013).

Ni, J.-Q. et al. A Drosophila resource of transgenic RNAi lines for neurogenetics. Genetics 182, 1089–1100 (2009).

Nichols, C. D., Becnel, J. & Pandey, U. B. Methods to Assay Drosophila Behavior. J Vis Exp (2012) doi:10.3791/3795.

Nordin, M., Bergman, D., Halje, M., Engström, W. & Ward, A. Epigenetic regulation of the Igf2/H19 gene cluster. Cell Proliferation 47, 189–199 (2014).

Novodvorsky, P. et al. klf2ash317 Mutant Zebrafish Do Not Recapitulate Morpholino-Induced Vascular and Haematopoietic Phenotypes. PLoS ONE 10, e0141611 (2015).

Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 29, 11–17 (2004).

Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801 (1980).

Nyberg, K. G. & Machado, C. A. Comparative Expression Dynamics of Intergenic Long Noncoding RNAs in the Genus Drosophila. Genome Biol Evol 8, 1839–1858 (2016).

Okazaki, Y. et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573 (2002).

Oliver, P. L. et al. Disruption of Visc-2, a Brain-Expressed Conserved Long Noncoding RNA, Does Not Elicit an Overt Anatomical or Behavioral Phenotype. Cereb. Cortex 25, 3572–3585 (2015).

Orjalo, A. V. & Johansson, H. E. Stellaris® RNA Fluorescence In Situ Hybridization for the Simultaneous Detection of Immature and Mature Long Noncoding RNAs in Adherent Cells. Methods Mol. Biol. 1402, 119–134 (2016).

Owusu-Ansah, E. & Perrimon, N. Modeling metabolic homeostasis and nutrient sensing in Drosophila: implications for aging and metabolic diseases. Dis Model Mech 7, 343–350 (2014).

Pachnis, V., Brannan, C. I. & Tilghman, S. M. The structure and expression of a novel gene activated in early mouse embryogenesis. EMBO J. 7, 673–681 (1988).

Pal Bhadra, M., Bhadra, U., Kundu, J. & Birchler, J. A. Gene Expression Analysis of the Function of the Male-Specific Lethal Complex in Drosophila. Genetics 169, 2061–2074 (2005).

Panzitt, K. et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology 132, 330–342 (2007).

Pardue, M. L. & Gall, J. G. MOLECULAR HYBRIDIZATION OF RADIOACTIVE DNA TO THE DNA OF CYTOLOGICAL PREPARATIONS. Proc Natl Acad Sci U S A 64, 600–604 (1969).

Parisien, M., Wang, X. & Pan, T. Diversity of human tRNA genes from the 1000-genomes project. RNA Biol 10, 1853–1867 (2013).

Park, C., Yu, N., Choi, I., Kim, W. & Lee, S. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. Bioinformatics 30, 2480–2485 (2014).

Park, C., Yu, N., Choi, I., Kim, W. & Lee, S. lncRNAtor: a comprehensive resource for

functional investigation of long non-coding RNAs. Bioinformatics 30, 2480–2485 (2014).

Park, E. & Maquat, L. E. Staufen-mediated mRNA decay. Wiley Interdiscip Rev RNA 4, 423–435 (2013).

Parker, N. F. & Shingleton, A. W. The coordination of growth among Drosophila organs in response to localized growth-perturbation. Dev. Biol. 357, 318–325 (2011).

Patil, D. P. et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature 537, 369–373 (2016).

Pauli, A., Rinn, J. L. & Schier, A. F. Non-coding RNAs as regulators of embryogenesis. Nat. Rev. Genet. 12, 136–149 (2011).

Pease, B., Borges, A. C. & Bender, W. Noncoding RNAs of the Ultrabithorax domain of the Drosophila bithorax complex. Genetics 195, 1253–1264 (2013).

Peng, J. et al. Expression profile analysis of long noncoding RNA in ER-positive subtype breast cancer using microarray technique and bioinformatics. Cancer Manag Res 9, 891–901 (2017).

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. Nature 379, 131–137 (1996).

Perkins, L. A. et al. The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics 201, 843–852 (2015).

Peter, I. S. & Davidson, E. H. Evolution of Gene Regulatory Networks that Control Embryonic Development of the Body Plan. Cell 144, 970–985 (2011).

Petruk, S. et al. Transcriptional elongation of non-coding bxd RNAs promoted by the Trithorax TAC1 complex represses Ubx by a transcriptional interference mechanism. Cell 127, 1209–1221 (2006).

Picao-Osorio, J., Lago-Baldaia, I., Patraquim, P. & Alonso, C. R. Pervasive Behavioral Effects of MicroRNA Regulation in Drosophila. Genetics 206, 1535–1548 (2017).

Pintacuda, G. et al. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol. Cell 68, 955-969.e10 (2017).

Place, E. S. & Smith, J. C. Zebrafish atoh8 mutants do not recapitulate morpholino phenotypes. PLoS ONE 12, e0171143 (2017).

Podbevšek, P. et al. Structural determinants of the SINE B2 element embedded in the long non-coding RNA activator of translation AS Uchl1. Sci Rep 8, 3189 (2018).

Poirier, F. et al. The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. Development 113, 1105–1114 (1991).

Polesello, C., Roch, F., Gobert, V., Haenlin, M. & Waltzer, L. Modeling cancers in Drosophila. Prog Mol Biol Transl Sci 100, 51–82 (2011).

Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038 (2010).

Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet. 5, e1000617 (2009).

Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. Cell 136, 629–641 (2009).

Port, F. & Bullock, S. L. Augmenting CRISPR applications in Drosophila with tRNA-flanked Cas9 and Cpf1 sgRNAs. Nat Methods 13, 852–854 (2016).

Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A 111, E2967–E2976 (2014).

Prasanth, K. V. & Spector, D. L. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev. 21, 11–42 (2007).

Qu, L. et al. Exosome-Transmitted lncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. Cancer Cell 29, 653–668 (2016).

Quinn, J. J. & Chang, H. Y. In situ dissection of RNA functional subunits by domain-specific chromatin isolation by RNA purification (dChIRP). Methods Mol. Biol. 1262, 199–213 (2015).

Quinn, J. J. et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat. Biotechnol. 32, 933–940 (2014).

Qureshi, I. A., Mattick, J. S. & Mehler, M. F. Long non-coding RNAs in nervous system function and disease. Brain Res 1338C, 20–35 (2010).

Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283 (2011).

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5, 877–879 (2008).

Ran, M. et al. Systematic Identification of Long Noncoding RNAs in Immature and Mature Porcine Testes. Biol. Reprod. 94, 77 (2016).

Rancourt, R. C., Harris, H. R., Barault, L. & Michels, K. B. The prevalence of loss of imprinting of H19 and IGF2 at birth. FASEB J 27, 3335–3343 (2013).

Rapicavoli, N. A. et al. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. eLife 2, (2013).

Rapicavoli, N. A., Poth, E. M. & Blackshaw, S. The long noncoding RNA RNCR2

directs mouse retinal cell specification. BMC Dev. Biol. 10, 49 (2010).

Raveendra, B. L. et al. Long noncoding RNA GM12371 acts as a transcriptional regulator of synapse function. Proc. Natl. Acad. Sci. U.S.A. 115, E10197–E10205 (2018).

Ravel-Chapuis, A. et al. The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. J Cell Biol 196, 699–712 (2012).

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. Genome Res. 11, 1114–1125 (2001).

Ren, X. et al. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep 9, 1151–1162 (2014).

Rinn, J. L. lncRNAs: Linking RNA to Chromatin. Cold Spring Harb Perspect Biol 6, a018614 (2014).

Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166 (2012).

Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–1323 (2007).

Rivas, F. V. et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. 12, 340–349 (2005).

Rudrapatna, V. A., Cagan, R. L. & Das, T. K. Drosophila cancer models. Dev. Dyn. 241, 107–118 (2012).

Ryder, E. et al. The DrosDel deletion collection: a Drosophila genomewide chromosomal deficiency resource. Genetics 177, 615–629 (2007).

Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development 128, 1275–1286 (2001).

Saito, K. et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 20, 2214–2222 (2006).

Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? Cell 146, 353–358 (2011).

Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F. & Ochiya, T. Emerging roles of long non-coding RNA in cancer. Cancer Sci. 109, 2093–2100 (2018).

Sauvageau, M. et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife 2, (2013).

Scadden, D. A NEAT Way of Regulating Nuclear Export of mRNAs. Molecular Cell 35, 395–396 (2009).

Schaefer, K. A. et al. Unexpected mutations after CRISPR-Cas9 editing in vivo. Nat. Methods 14, 547–548 (2017).

Scheper, G. C. et al. Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. Nature Genetics 39, 534–539 (2007).

Scheper, G. C., van der Knaap, M. S. & Proud, C. G. Translation matters: protein synthesis defects in inherited disease. Nature Reviews Genetics 8, 711–723 (2007).

Scherrer, K., Latham, H. & Darnell, J. E. DEMONSTRATION OF AN UNSTABLE RNA AND OF A PRECURSOR TO RIBOSOMAL RNA IN HELA CELLS*. Proc Natl Acad Sci U S A 49, 240–248 (1963).

Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. Cancer Cell 29, 452–463 (2016).

Schmitz, S. U., Grote, P. & Herrmann, B. G. Mechanisms of long noncoding RNA function in development and disease. Cell. Mol. Life Sci. 73, 2491–2509 (2016).

Schöck, F. & Perrimon, N. Cellular processes associated with germ band retraction in Drosophila. Dev. Biol. 248, 29–39 (2002).

Schulz, D. et al. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. Cell Syst 6, 25-36.e5 (2018).

Serviss, J. T., Johnsson, P. & Grandér, D. An emerging role for long non-coding RNAs in cancer metastasis. Front Genet 5, (2014).

Sharp, P. A. The discovery of split genes and RNA splicing. Trends Biochem. Sci. 30, 279–281 (2005).

Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116–120 (2012).

Sheng, L., Ye, L., Zhang, D., Cawthorn, W. P. & Xu, B. New Insights Into the Long Non-coding RNA SRA: Physiological Functions and Mechanisms of Action. Front Med (Lausanne) 5, (2018).

Shukla, C. J. et al. High-throughput identification of RNA nuclear enrichment sequences. EMBO J. 37, (2018).

Sigova, A. A. et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. U.S.A. 110, 2876–2881 (2013).

Simon, M. D. et al. The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. U.S.A. 108, 20497–20502 (2011).

Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. Nature 440, 448–455 (2006).

Singh, A. K. & Lakhotia, S. C. Dynamics of hnRNPs and omega speckles in normal and heat shocked live cell nuclei of Drosophila melanogaster. Chromosoma 124, 367–383 (2015).

Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415, 810–813 (2002).

Soeiro, E., Birnboim, H. C. & Darnell, J. E. Rapidly labeled HeLa cell nuclear RNA: II. Base composition and cellular localization of a heterogeneous RNA fraction. Journal of Molecular Biology 19, 362–372 (1966).

Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437 (2004).

Song, M. S. et al. Nuclear PTEN regulates the APC-CDH1 tumor-suppressive complex in a phosphatase-independent manner. Cell 144, 187–199 (2011).

Song, X., Wang, X., Arai, S. & Kurokawa, R. Promoter-associated noncoding RNA from the CCND1 promoter. Methods Mol. Biol. 809, 609–622 (2012).

Soshnev, A. A. et al. A conserved long noncoding RNA affects sleep behavior in Drosophila. Genetics 189, 455–468 (2011).

Spadaro, P. A. et al. Long Noncoding RNA-Directed Epigenetic Regulation of Gene Expression Is Associated With Anxiety-like Behavior in Mice. Biol. Psychiatry 78, 848–859 (2015).

Spadaro, P. A. et al. Long Noncoding RNA-Directed Epigenetic Regulation of Gene Expression Is Associated With Anxiety-like Behavior in Mice. Biol. Psychiatry 78, 848–859 (2015).

St Johnston, D. The art and design of genetic screens: Drosophila melanogaster. Nat. Rev. Genet. 3, 176–188 (2002).

St Laurent, G., Wahlestedt, C. & Kapranov, P. The Landscape of long non-coding RNA classification. Trends Genet 31, 239–251 (2015).

Stefani, G. & Slack, F. J. Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. 9, 219–230 (2008).

Stuckenholz, C., Meller, V. H. & Kuroda, M. I. Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. Genetics 164, 1003–1014 (2003).

Sun, M. & Kraus, W. L. Minireview: Long noncoding RNAs: new 'links' between gene expression and cellular outcomes in endocrinology. Mol. Endocrinol. 27, 1390–1402 (2013).

Sun, M. & Kraus, W. L. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. Endocr. Rev. 36, 25–64 (2015).

Sun, S. et al. Jpx RNA Activates Xist by Evicting CTCF. Cell 153, 1537–1551 (2013).

Sun, X. & Heckscher, E. S. Using Linear Agarose Channels to Study Drosophila Larval Crawling Behavior. J Vis Exp (2016) doi:10.3791/54892.

Sun, Y. et al. Long non-coding RNA HOTTIP promotes BCL-2 expression and induces chemoresistance in small cell lung cancer by sponging miR-216a. Cell Death & Disease 9, 85 (2018).

Sun, Y. et al. A long non-coding RNA HOTTIP expression is associated with disease progression and predicts outcome in small cell lung cancer patients. Molecular Cancer 16, 162 (2017).

Sweta, S., Dudnakova, T., Sudheer, S., Baker, A. H. & Bhushan, R. Importance of Long Non-coding RNAs in the Development and Disease of Skeletal Muscle and Cardiovascular Lineages. Front Cell Dev Biol 7, 228 (2019).

Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462, 799–802 (2009).

Tang, S.-S., Zheng, B.-Y. & Xiong, X.-D. LincRNA-p21: Implications in Human Diseases. Int J Mol Sci 16, 18732–18740 (2015).

Tano, K. & Akimitsu, N. Long non-coding RNAs in cancer progression. Front. Genet. 3, (2012).

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

The RNAcentral Consortium et al. RNAcentral: a comprehensive database of non-coding RNA sequences. Nucleic Acids Res. 45, D128–D134 (2017).

The RNAcentral Consortium et al. RNAcentral: a comprehensive database of non-coding RNA sequences. Nucleic Acids Res. 45, D128–D134 (2017).

Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell 143, 390–403 (2010).

Tollervey, D. & Kiss, T. Function and synthesis of small nucleolar RNAs. Curr. Opin. Cell Biol. 9, 337–342 (1997).

Tran, U. M. et al. LincRNA-p21 acts as a mediator of ING1b-induced apoptosis. Cell Death Dis 6, e1668 (2015).

Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).

Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell 39, 925–938 (2010).

Tripathi, V. et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet. 9, e1003368 (2013).

Tsai, M.-C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689–693 (2010).

Tsoi, L. C. et al. Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. Genome Biol. 16, 24 (2015).

Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. 15, 342–348 (2005).

Turner, M., Galloway, A. & Vigorito, E. Noncoding RNA and its associated proteins as regulatory elements of the immune system. Nat Immunol 15, 484–491 (2014).

Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat. Rev. Genet. 17, 601–614 (2016).

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147, 1537–1550 (2011).

Vagin, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313, 320–324 (2006).

Valenzuela, D. M. et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat. Biotechnol. 21, 652–659 (2003).

Venken, K. J. T. & Bellen, H. J. Transgenesis upgrades for Drosophila melanogaster. Development 134, 3571–3584 (2007).

Venken, K. J. T. et al. Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat. Methods 6, 431–434 (2009).

Venter, J. C. et al. The sequence of the human genome. Science 291, 1304–1351 (2001).

Vickers, T. A. et al. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. J. Biol. Chem. 278, 7108–7118 (2003).

Volders, P.-J. et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 41, D246-251 (2013).

Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718 (2009).

Wallace, B. Studies on irradiated populations of Drosophila melanogaster. J Genet 54, 280–293 (1956).

Wang, A. H. et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature 282, 680–686 (1979).

Wang, A., Wang, J., Liu, Y. & Zhou, Y. Mechanisms of Long Non-Coding RNAs in the Assembly and Plasticity of Neural Circuitry. Front Neural Circuits 11, (2017).

Wang, D. O. et al. A quick and simple FISH protocol with hybridization-sensitive fluorescent linear oligodeoxynucleotide probes. RNA 18, 166–175 (2012).

Wang, H. et al. Dendritic BC1 RNA in translational control mechanisms. The Journal of Cell Biology 171, 811–821 (2005).

Wang, J. et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 38, 5366–5383 (2010).

Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. Mol. Cell 43, 904–914 (2011).

Wang, K. C. et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–124 (2011).

Wang, W., Brunet, F. G., Nevo, E. & Long, M. Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster. PNAS 99, 4448–4453 (2002).

Wang, X. et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454, 126–130 (2008).

Wang, Y.-F., Zhang, S., Li, X.-Q. & Wang, Y. Expression of lncRNA HULC in cervical cancer and its correlation with tumor progression and patient survival. Eur Rev Med Pharmacol Sci 20, 3987–3991 (2016).

Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737 (1953).

Wei, F. et al. Cloning and characterization of the putative AFAP1-AS1 promoter region. J Cancer 10, 1145–1153 (2019).

Wen, J. et al. Diversity of miRNAs, siRNAs, and piRNAs across 25 Drosophila cell lines. Genome Res. 24, 1236–1250 (2014).

Wen, K. et al. Critical roles of long noncoding RNAs in Drosophila spermatogenesis. Genome Res. 26, 1233–1244 (2016).

Weng, R. & Cohen, S. M. Drosophila miR-124 regulates neuroblast proliferation through its target anachronism. Development 139, 1427–1434 (2012).

West, J. A. et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell 55, 791–802 (2014).

Wilk, R., Hu, J., Blotsky, D. & Krause, H. M. Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. Genes Dev. 30, 594–609 (2016).

Wilkin, F. et al. H19 sense and antisense transgenes modify insulin-like growth factor-II mRNA levels. Eur. J. Biochem. 267, 4020–4027 (2000).

Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb Perspect Biol 3, (2011).

Wilusz, C. J. & Wilusz, J. HuR and translation--the missing linc(RNA). Mol. Cell 47, 495–496 (2012).

Winkler, W. C. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Curr Opin Chem Biol 9, 594–602 (2005).

Winzi, M. et al. The long noncoding RNA lncR492 inhibits neural differentiation of murine embryonic stem cells. PLOS ONE 13, e0191682 (2018).

Wixon, J. & O'Kane, C. Featured organism: Drosophila melanogaster. Yeast 17, 146–153 (2000).

Wu, L. & Belasco, J. G. Examining the influence of microRNAs on translation efficiency and on mRNA deadenylation and decay. Meth. Enzymol. 449, 373–393 (2008).

Wu, Y. et al. Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, Bombyx mori. PLOS ONE 11, e0147147 (2016).

Xing, Z. et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell 159, 1110–1125 (2014).

Xue, Y. et al. Genetic variants in lncRNA HOTAIR are associated with risk of colorectal cancer. Mutagenesis 30, 303–310 (2015).

Yamashita, A., Shichino, Y. & Yamamoto, M. The long non-coding RNA world in yeasts. Biochim. Biophys. Acta 1859, 147–154 (2016).

Yan, R. et al. Genetic variants in lncRNA SRA and risk of breast cancer. Oncotarget 7, 22486–22496 (2016).

Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. Cell 165, 976–989 (2016).

Yanaihara, N. et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 9, 189–198 (2006).

Yang, W. et al. MiR-146b-5p overexpression attenuates stemness and radioresistance of glioma stem cells by targeting HuR/lincRNA-p21/β-catenin pathway. Oncotarget 7, 41505–41526 (2016).

Yang, Y.-L. et al. Early growth response protein-1 upregulates long noncoding RNA Arid2-IR to promote extracellular matrix production in diabetic kidney disease. Am. J. Physiol., Cell Physiol. 316, C340–C352 (2019).

Yang, Y., Wen, L. & Zhu, H. Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein. Cell Biosci 5, (2015).

Yang, Z. et al. HULC and H19 Played Different Roles in Overall and Disease-Free Survival from Hepatocellular Carcinoma after Curative Hepatectomy: A Preliminary Analysis from Gene Expression Omnibus. Dis Markers 2015, (2015).

Yeasmin, F., Yada, T. & Akimitsu, N. Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics. Front Genet 9, (2018).

Yildirim, E. et al. Xist RNA is a potent suppressor of hematologic cancer in mice. Cell 152, 727–742 (2013).

Yoon, J.-H., Abdelmohsen, K. & Gorospe, M. Posttranscriptional gene regulation by long noncoding RNA. J. Mol. Biol. 425, 3723–3730 (2013).

Yoon, J.-H. et al. Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. Nat Commun 4, 2939 (2013).

Yoon, J.-H. et al. LincRNA-p21 suppresses target mRNA translation. Mol. Cell 47, 648–655 (2012).

Young, R. S. et al. Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome. Genome Biol Evol 4, 427–442 (2012).

Yu, Z. et al. Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the Drosophila genome. Biol Open 3, 271–280 (2014).

Yu, Z. et al. Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics 195, 289–291 (2013).

Yuan, L. et al. The Drosophila neurogenin Tap functionally interacts with the Wnt-PCP pathway to regulate neuronal extension and guidance. Development 143, 2760–2766 (2016).

Yue, Y. et al. Integrated Analysis of the Roles of Long Noncoding RNA and Coding RNA Expression in Sheep (Ovis aries) Skin during Initiation of Secondary Hair Follicle. PLoS One 11, (2016).

Zeng, Y. & Cullen, B. R. RNA interference in human cells is restricted to the cytoplasm. RNA 8, 855–860 (2002).

Zhang, B. et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep 2, 111–123 (2012).

Zhang, D. et al. Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. Cell Research 19, 429–438 (2009).

Zhang, W. & Cohen, S. M. The Hippo pathway acts via p53 and microRNAs to control proliferation and proapoptotic gene expression during tissue growth. Biol Open 2, 822–828 (2013).

Zhao, L., Liu, L., Wang, S., Wang, H. & Jiang, J. Transcriptome profiles of metamorphosis in the ornamented pygmy frog Microhyla fissipes clarify the functions of thyroid hormone receptors in metamorphosis. Sci Rep 6, (2016).

Zhao, X. et al. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nature Communications 9, 5056 (2018).

Zhao, Y. et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res 44, D203–D208 (2016).

Zhong, J. et al. BC1 regulation of metabotropic glutamate receptor-mediated neuronal excitability. J. Neurosci. 29, 9977–9986 (2009).

Zhou, Q., Huang, X. R., Yu, J., Yu, X. & Lan, H. Y. Long Noncoding RNA Arid2-IR Is a Novel Therapeutic Target for Renal Inflammation. Mol. Ther. 23, 1034–1043 (2015).

Zhou, S. et al. Systematical analysis of lncRNA-mRNA competing endogenous RNA network in breast cancer subtypes. Breast Cancer Res. Treat. 169, 267–275 (2018).

Zong, X. et al. Knockdown of nuclear-retained long noncoding RNAs using modified DNA antisense oligonucleotides. Methods Mol. Biol. 1262, 321–331 (2015).

Zou, S., Meadows, S., Sharp, L., Jan, L. Y. & Jan, Y. N. Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. PNAS 97, 13726–13731 (2000).

Zou, T., Wang, P. L., Gao, Y. & Liang, W. T. Long noncoding RNA HOTTIP is a significant indicator of ovarian cancer prognosis and enhances cell proliferation and invasion. Cancer Biomark (2018) doi:10.3233/CBM-181727.

Zuo, L. et al. Long non-coding RNAs in psychiatric disorders. Psychiatr Genet 26, 109–116 (2016).

Züst, R. et al. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat. Immunol. 12, 137–143 (2011).